P0001: Gene Editing/CRISPR

Live Cell CRISPR-Imaging in Plants

Steven Dreissig[1], Simon Schiml[2], Patrick Schindele[2], Oda Weiss[1], Twan Rutten[3], Veit Schubert[3], Evgeny Gladilin[1], Florian Mette[4], Holger Puchta[5] and **Andreas Houben**[3], (1)IPK, Gatersleben, Germany, (2)KIT, Karlsruhe, Germany, (3)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (4)King Abdullah University of Science & Technology, Thuwal, Saudi Arabia, (5)Karlsruhe Institute of Technology, Karlsruhe, Germany

Elucidating the spatio-temporal organization of the genome inside the nucleus is imperative to understand the regulation of genes and non-coding sequences during development and environmental changes. Emerging techniques of chromatin imaging promise to bridge the long-standing gap between sequencing studies which reveal genomic information and imaging studies that provide spatial and temporal information of defined genomic regions. Here, we demonstrate such an imaging technique based on two orthologues of the bacterial CRISPR-Cas9 system. By fusing eGFP/mRuby2 to the catalytically inactive version of *Streptococcus pyogenes* and *Staphylococcus aureus* Cas9, we show robust visualization of telomere repeats in live leaf cells of *Nicotiana benthamiana*. By tracking the dynamics of telomeres visualized by CRISPR-dCas9, we reveal long range telomere movements of up to 2 µm within 30 minutes during interphase. Furthermore, we show that CRISPR-dCas9 can be combined with fluorescence-labelled proteins to visualize DNA-protein interactions *in vivo*. By simultaneously using two dCas9 orthologues, we pave the way for imaging of multiple genomic loci in live plants cells. CRISPR-imaging bears the potential to significantly improve our understanding of the dynamics of chromosomes in live plant cells.

P0002: Gene Editing/CRISPR

CRISPRdirect & GGGenome: Web-Based Software for CRISPR-Cas9 Guide RNA Design with Fast and Sensitive Off-Target Searches

Yuki Naito and Hidemasa Bono, Database Center for Life Science, Mishima, Japan

CRISPRdirect (http://crispr.dbcls.jp/) is a simple and functional web-based online software for designing rational CRISPR-Cas9 guide RNAs. The software selects highly specific gRNAs by performing searches against entire genome using GGGenome (http://GGGenome.dbcls.jp/) nucleotide sequence search software and k-mer hash tables. GGGenome quickly searches short sequences from various kind of genomic sequences allowing mismatches and gaps. The query sequences may contain degenerate nucleotide characters (e.g. N, R, Y) which typically appear in PAM sequences. The recent updates of CRISPRdirect and GGGenome include support of more than 350 organisms including plants, animals and fungi. We also consider adding another species if their genomic sequences are publicly available. These tools also provide REST API which is useful for processing large number of sequences in an automated manner. All services of GGGenome and CRISPRdirect web servers are freely available to all users.

P0003: Gene Editing/CRISPR

Development of a CRISPR/Cas9 Large DNA Fragment Targeting Technique for Plant Genomes

Céline Chantry-Darmon[1], **Carine Satgé**[1], Nicolas Milon[2], Arnaud Bellec[1], William Marande[1], Stephane Cauet[1], Caroline Callot[1], Pauline Duriez[3], Stéphane Munos[3], Aurélien Bancaud[2] and Helene Berges[1], (1)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (2)CNRS, LAAS, 7 avenue du colonel Roche, 31400 Toulouse, France, (3)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, TOULOUSE, France

In the context of a need of improvement and adaptation of plants, genomic exploration is one of the strategic approaches of choice. This knowledge facilitates the identification of genes, which can play a role in biotic or abiotic resistance, in yield or in quality process. However, the exploration of plant genomes remains challenging due to the complexity of plant genomes in terms of size, repetitive elements content and various levels of ploidy.

Moreover, due to high intra-species variability, a quality reference sequence is not enough. A precise and reliable information of a genomic region linked to a trait of interest in a specific genotype is requested to understand complex biological processes. However, it could still be very challenging to identify precisely the specific gene or region responsible for the agronomic trait of interest. New strategies for efficiently targeting large regions of interest in complex genomes are really needed to be able to link a phenotype to a genotype.

Here, we report a new sequence capture approach for large DNA fragments in eukaryote genomes based on the programmable endonuclease function of the CRISPR/Cas9 system. We improved and adapted the first steps of the CATCH method (Cas9-Assisted Targeting of CHromosomal segments as described by Jiang et al., 2015) to cut a specific locus in *Medicago truncatula*. We will also present the development of this approach on a larger and more complex plant genome region in Sunflower.

P0004: Gene Editing/CRISPR

Application of CRISPR/Cpf1-Based Genome Editing in Polyploid Wheat

Wei Wang[1], Qianli Pan[1], Fei He[1], Monica Fernandez de Soto[1,2], Jie Ren[1,2], Yueying Chen[1], Bin Tian[1], Alina Akhunova[1,2], Shiaoman Chao[3], Harold N. Trick[1] and Eduard Akhunov[1], (1)Kansas State University, Department of Plant Pathology, Manhattan, KS, (2)Integrated Genomics Facility, Manhattan, KS, (3)USDA-ARS, Fargo, ND

CRISPR/Cas9 has been widely applied in many organisms as a powerful genome editing tool. However, its target sites amenable to editing are limited by the 3'-end NGG proto-spacer adjacent motif (PAM). The CRISPR/Cpf1 requires 5'-end TTV or TTTV PAMs providing opportunities for targeting the AT-rich regions. The genome editing ability of FnCpf1 and LbCpf1 were assessed in wheat by combining transient expression in the wheat protoplasts and next generation sequencing (NGS) of the target regions. While no genome editing events were found for FnCpf1, about 1/3 of the designed targets for LbCpf1 showed the evidence of genome editing. The efficiency of editing was up to 10%, comparable to that of SpCas9 in the wheat protoplasts. We are also testing the editing efficiency of the mutated LbCpf1 (G532R/K538V/Y542R, henceforth LbCpf1m), which in mammalian cells was shown can induce mutations with the 5'-end of the TATV PAM. Multiplex gene editing using Cas9 is somewhat limited by the size of the transgenic constructs. The ability of Cpf1 to process its own CRISPR RNA (crRNA) and the shorter length (43 nucleotides) of crRNA make it a promising multiplex genome editing tool. Multiplex gene editing constructs with different numbers of crRNAs under the control of a single promoter were constructed. All multiplex genome editing

constructs generated indels/insertions at the targets sites with the efficiency comparable to that of a single crRNA construct. Our results show that the LbCpf1 can further expand the set of tools available for engineering the wheat genome.

P0005: Gene Editing/CRISPR

CRISPR/CAS9 Mediated Down-Regulation of Lignin-Biosynthetic Gene Coumarate 3-Hydroxylase (C3H) in Rice (Oryza sativa spp. japonica cv. Nipponbare)

Bikram Poudel, Sathish Kumar Ponniah and Muthusamy Manoharan, University of Arkansas at Pine Bluff, Pine Bluff, AR

Rice straw is one of the most abundant lignocellulosic agricultural residue worldwide. It can be utilized to produce lignocellulosic biofuel due to the presence of polysaccharides like cellulose and hemicellulose. However, lignin physically impedes the enzymatic deconstruction of polysaccharides during conversion to bioethanol. Lignin is synthesized through phenylpropanoid pathway and composed of three main monolignols, the *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units. Coumarate 3-hydroxyase (C3H) is responsible for the synthesis of G and S subunits. Downregulation of C3H results in redirection of phenylpropanoid flux thereby increasing the amount of p-hydroxyphenyl (H) units and increase the synthesis of flavonoids. Here, we used CRISPR/Cas9 to downregulate C3H gene leading to reduced lignin content in rice straw. Three guide RNA (gRNA) were designed to target three sites of gene C3H1 (*LOC_ Os05g0494000*). The clone with pRGEB32 backbone was sequenced and then performed *Agrobacterium*-mediated transformation generating 24 T_0 independent events. PCR analysis of T_0 plants indicated the presence of M13/Ubiquitin, Cas9 and hygromycin resistance cassette. Initial sequencing result of T_0 plants revealed insertion, deletion and substitution mutations in and around the gRNA region. The mutant plants with Insertions or deletions (InDels) will be grown to obtain T_1 plants. The T_1 plants will be analyzed for segregation and desired gene modification. The mutant plants without Cas9 will then be analyzed for the lignin content as well as changes in flavonoid content.

P0006: Gene Editing/CRISPR

CRISPR/Cas9-Mediated Editing of Lignin Biosynthetic Genes, Hydroxycinnamoyl CoA: Shikimate Hydroxycinnamoyl Transferase (HCT) and Caffeic Acid O-Methyltransferase (COMT), in Rice (Oryza sativa L. ssp. japonica cv. Nipponbare)

Kumar Shrestha, Sathish Kumar Ponniah and Muthusamy Manoharan, University of Arkansas at Pine Bluff, Pine Bluff, AR

Rice straw is one of the largest biomasses in the world with a potential for biofuel production. However, lignin present in rice straw impedes the enzymatic hydrolysis of cellulose and hemicellulose, thus inhibiting cellulosic biofuel production. Here, our objective was to downregulate the lignin content of rice straw by editing the lignin biosynthetic genes: *hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase* (HCT) and *caffeic acid O-methyltransferase* (COMT) using CRISPR/Cas9. HCT drives the phenylpropanoid pathway towards the synthesis of S (Syringyl) and G (Guaiacyl) lignin, while COMT directs the synthesis of S-Lignin. Three gRNAs for HCT (*LOC_Os04g42250)* and two for COMT (*LOC_Os02g57760)* were designed to target respective genes for mutations. *Agrobacterium*-mediated transformation was used to develop T_0 plants: 21 and 26 independent events of HCT and COMT lines were developed respectively. PCR was carried out to confirm the presence of Cas9, hygromycin and gRNAs in T_0 plants, which indicated the transgenic nature of regenerated HCT and COMT lines. The preliminary sequencing result of positive T_0 plants revealed various insertions and deletions (InDels) and substitutions in and around target sites. Further, T_1 plants will be grown from the confirmed mutant plants and examined for desired mutations. The null segregants identified from the T_1 plants will then be analyzed for lignin content.

P0007: Gene Editing/CRISPR

CRISPR/Cas9-Mediated Gene Editing in Rice (Oryza sativa L. japonica cv. Katy) for Stable Resistance Against Blast Fungus (Magnaporthe oryzae)

Bed Prakash Bhatta[1], Sathish Kumar Ponniah[1], Yulin Jia[2] and Muthusamy Manoharan[1], (1)University of Arkansas at Pine Bluff, Pine Bluff, AR, (2)USDA-ARS, Stuttgart, AR

Rice blast is one of the recurring and devastating diseases in the USA and worldwide. In the USA, the blast-resistance (*R*) genes present in a tropical japonica cultivar Katy reduces blast damages from 1990 to present. The cultivar is used as a principal donor of blast *R* genes in developing numerous elite US resistant cultivars. The objective of this research is to identify the role of *R* genes in blast resistance of Katy using CRISPR/Cas9 targeting a nucleotide binding site and leucine rich repeat containing gene in rice (*LOC_Os12g18374*) at the *Pi-ta*/*Pita2*/*Ptr* region. Katy with the *Pi-ta/Pita2/Ptr* gene cluster is resistant to rice blast isolates that contain the corresponding avirulence gene *AVR-Pita* except a virulent strain *IE-1k* that does not contain *AVR-Pita*. Two vectors containing single guide RNA (S1) encoding alanine and combination of three gRNAs (S2) encoding asparagine, serine and alanine were used to induce double-strand breaks at the intended target sites in Katy. Using *Agrobacterium*-mediated transformation, 24 independent events of Katy were generated in each S1 and S2 transformations. Genomic DNA from putative mutant T_0 plants was extracted and analyzed by PCR using M13/Ubiquitin, Cas9, and hygromycin primers. Results confirmed the presence of M13/Ubiquitin, Cas9 and hygromycin genes in T_0 plants. The genomic DNA from T_0 mutant plants was further sequenced using gene-specific primer upstream of the target locus. Series of indels and substitutions were observed in and around the gRNA sequence. T_1 plants will be grown to identify the null segregants and scored after blast infection.

P0008: Gene Editing/CRISPR

High Efficient Induction of Single DNA Conversion in Tomato by Target-Aid, the Base Editing CRISPR-Cas9 System

Sachiko Kashojiya[1], Yumi Heta[1], Kenji Miura[1], Keiji Nishida[2], Akihiko Kondo[2], Hiroshi Ezura[3] and Tohru Ariizumi[3], (1)University of Tsukuba, Tsukuba, Japan, (2)Kobe University, Kobe, Japan, (3)University of Tsukuba, Ibaraki, Japan

Target-AID is the modified CRISPR-Cas9 system fused with the activation-induced cytidine deaminase (PmCDA1) derived from sea lamprey and allows for single base substitution within the targeted region specified by sgRNA. Target-AID uses either nuclease deficient dead Cas9 (dCas9) or nickase Cas9 (nCas9) for fusion with CDA1 and previously appeared to induce efficient DNA insertion, deletion and substitution in yeast, mammalian cells and plants.

This is study reports the application of Target-AID to tomato by targeting the *SlDELLA* gene which encodes a negative regulator of phytohormone gibberellin signaling and examined the spectrum of mutations caused by the system. We constructed conventional CRISPR-

Cas9 (*Cas9*), *nCas9-PmCDA1* and *dCas9-PmCDA1* as Target-AID, and *nCas9* (without fusion) as a negative control and analyzed targeted sequences in T_0 and T_1 plants. *Cas9* induced insertion and deletion in both generations as expected. In contrast, *nCas9-PmCDA1* induced insertion, deletion and substitutions, and six different patterns of DNA conversion were found by this system. Further, we confirmed two mutation spectrums (CAC to tAt or CAt) resulted in increased GA sensitivity including efficient parthenocarpy. However, both nCas9 and *dCas9-PmCDA1* rarely induced mutations within sgRNA sequence in both T_0 and T_1 plants. Taken together, it was demonstrated that *nCas9-PmCDA1* could be powerful tool to modify crop traits by generating novel single base polymorphisms.
This work was supported by Cabinet Office, Government of Japan, Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for creating next-generation agriculture, forestry and fisheries" (funding agency: Bio-oriented Technology Research Advancement Institution, □□□□).

P0009: Gene Editing/CRISPR

Trial for Generating Parthenocarpic Tomato by the CRISPR/Cas9 System Employed Base Editing Enzyme

Tsubasa Narihiro[1], Kohsuke Kuwabara[1], Keiji Nishida[2], Akihiko Kondo[2], Hiroshi Ezura[3] and Tohru Ariizumi[3], (1)University of Tsukuba, Tsukuba, Japan, (2)Kobe University, Kobe, Japan, (3)University of Tsukuba, Ibaraki, Japan

Parthenocarpy is the natural or artificial induced fruit formation without pollination or pollination in fruit bearing crops and is one of the valuable traits for tomato breeding since it could improve fruit set efficiency even under adverse environmental condition such as high temperature and omit cumbersome pollination works. However currently known parthenocarpic genetic loci are known to be associated with inferior characteristics such as earlier softening and cracking. This study aims to generate new parthenocarpic cultivar which overcomes these drawback traits by targeting naturally found parthenocarpy genetic loci with the previously developed modified CRISPR/Cas9 system employing the cytidine deaminase, a base editing enzyme that catalyzes the irreversible hydrolytic deamination of cytidine to uridine. This modified CRISPR/Cas9 system, called Target-AID (activation induced cytidine deaminase), has previously demonstrated to work as targeted mtuagenesis approach including efficient DNA deletion, insertion and conversion within the DNA sequence specified by sgRNA. We have so far obtained 60 transgenic tomato plants in the genetic background of dwarf cultivar Micro-Tom and confirmed targeted mutagenesis in their offspring T1 plants. In this presentation, we show results of currently ongoing genotyping analysis and their phenotypic characterization.
This work was supported by Cabinet Office, Government of Japan, Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for creating next-generation agriculture, forestry and fisheries"(funding agency: Bio-oriented Technology Research Advancement Institution, □□□□)

P0010: Gene Editing/CRISPR

Breaking Self-Incompatibility in Diploid Potato using the CRISPR/Cas9 System

Felix E. Enciso[1], Daniel Zarka[1], Satya Swathi Nadakuduti[1], C. Robin Buell[2] and David Douches[1], (1)Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, (2)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

Potato breeding programs rely on the discovery and introgression of genes of interest. However, the polyploid nature of potatoes hampers the fixation of desirable alleles in new cultivars. Therefore, creating inbred diploid potatoes represents an alternative strategy for obtaining homozygous lines. Diploid potatoes possess self-incompatibilty (SI) that forces outcrossing resulting in heterozygosity in majority of alleles, limiting inbred line development. In the gametophytic SI system, the S-locus F-box protein expressed in the pollen does not recognize its own S haplotype of the style (S-RNAse), expressed in the pistil, producing cytotoxic effects that inhibit the elongation of self-pollen tubes. The aim of this project is to generate a targeted knock-out of the S-RNase locus in SI potato lines using Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated systems (CRISPR/Cas9) technology in an effort to avoid self-pollen toxicity. Three diploid self-incompatible lines from the Michigan State University potato breeding program were used. Seven reported S-RNAse genes were selected according to their functional evidence in SI mechanism and compared against the parental genomes. The selected and functionally verified candidate gene was amplified. Two gRNAs were designed from conserved regions on each predicted exon and assembled in an expression vector carrying the Cas9, U6 promoter and scaffold guide RNA. The expression cassette was inserted into the T-DNA region *Agrobacterium tumefaciens*. Agrobacterium-mediated transformation using stem/leaf explants was performed. Regeneration events obtained from selection media will be further evaluated using T7 endonuclease assay and sequencing. The S-RNase knock out mutants will be verified for self-compatibility and S-RNase expression.

P0011: Gene Editing/CRISPR

Protoplast-Based CRISPR/Cas9 for Modification of Petunia Flower Color

Geung-Joo Lee, Luhua Tu and Saminathan Subburaj, Chungnam National University, Daejeon, South Korea

Recent advances in genome information and new technologies such as genome editing allow better understanding of complicated genetic phenomena in many horticultural crops. An *F3H (flavanone 3-hydroxylase*) gene involved in flavonoid biosynthetic pathway has been recognized as one of the major encoding components of floral pigmentation in petunia belonging to Solanaceae family. A CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) system has been reported to have target-specific modifying potential in many crops. To investigate the function of *F3H* in flower color variation by a gene-directed mutagenesis, we delivered directly a purified Cas9 protein and single guide RNA (sgRNA) into isolated protoplast cells of Petunia (cv. 'midnight'). After transient introduction of RNA-guided endonuclease (RGEN) ribonucleoproteins (RNPs) complex with different sgRNAs targeting five *F3H* regions, mutagenesis at the targeted loci was detected by HRM assay and confirmed by targeted deep sequencing. Wide ranges of INDEL (deletions and insertion types) were observed in the cloned nucleotide sequences of *F3H* from those protoplast transfectants, but not in putative off-target sites. Targeted deep DNA sequencing analysis revealed higher mutation rates from those five F3H-RGEN target sites, ranging from 0.8% to 49.3% with an average of 20.8 ± 7.2 %. We will further provide the average ratio of deletion to insertion produced collectively by the five F3H-RGEN target sites (F3H_1 to 5). The current results indicated that direct delivery of GEN RNPs into protoplast cells without the use of transgenic approaches can be exploited as an efficient tool for genetic manipulation of flower colors in ornamental petunia, which is an economically important crop, but with ralatively inadequate genome information.

P0012: Gene Editing/CRISPR

Utilizing CRISPR Genome Editing to Rapidly Domesticate the Winter Annual Oilseed Crop Pennycress (*Thlaspi arvense*)

Malihe Esfahanian, Michaela McGinn, Brice Jarvis, Taylor Suo and **John C. Sedbrook**, Illinois State University, Normal, IL

Pennycress (*Thlapsi arvense*) is an emerging oilseed crop closely related to Arabidopsis and rapeseed canola that holds considerable agronomic and economic potential in producing seed oil to be used as a liquid biofuels feedstock. Pennycress possesses a unique combination of attributes including extreme cold tolerance, rapid growth, over-wintering growth habit, and a natural ability to produce copious amounts of seeds high in oil and protein. Pennycress could generate billions of liters of oil annually throughout temperate regions of the world without displacing food crops or requiring land use changes. For example, pennycress can be grown throughout the 40 million-acre U.S. Midwest Corn Belt during the fall through spring months, double-cropped between corn and soybeans on otherwise fallow farmland. Post oil extraction, the pennycress seed meal can be used as a high protein, nutrient-filled animal feed. Being that current pennycress varieties are not far removed from wild strains, we are working to rapidly improve breeding-line agronomic traits such as seed dormancy, pod shatter, seed oil and meal quality, and time to maturity, by using both forward and reverse genetics approaches. This poster will highlight our efforts in using CRISPR genome editing tools to rapidly improve pennycress as a profitable oilseed-producing winter cover crop, employing knowledge gained from decades of research on Arabidopsis and other Brassicaceae.

P0013: Gene Editing/CRISPR

Current Opportunities for Accelerated Livestock Breeding Using CRISPR-Cas9

Vrushali Patil[1], Simon Lillico[1], Chris Proudfoot[1], John Hickey[2] and Bruce Whitelaw[1], (1)The Roslin Institute, Midlothian, United Kingdom, (2)The Roslin Institute, Edinburgh, United Kingdom

Genetic engineering has seen unparalleled success in the last decade with the advent of CRISPR-Cas9 technology. Due to the demands of food supply around the world, accelerated improvements in agricultural productivity are needed if the mounting human population has to be sustained on this planet. Desirable traits are often complex and regulated by multiple genes hence it becomes time consuming to unravel how a particular gene influences a trait and how we can exploit it to produce the ideal phenotype. Precise gene editing – enabling indel production, sequence deletion and SNP introgression – offers the opportunity to induce targeted alterations in a genome sequence which can be then reliably quantified, thus adding greatly to our knowledge of gene function and its associated trait. This technology has been successfully applied to provide insights into reproductive efficiency, neurological stability and even resistance to deadly pathogens. The simplicity of CRISPR-Cas9 has also allowed us to introduce desirable mutations directly into the target species, saving considerable time and effort. We present our current projects aiming to enhance livestock breeding and consequently, our society.

P0014: Gene Editing/CRISPR

Phenotypic Effects of a Novel CRISPR/Cas9 Knockout of HMGA2 in Mice

Jingyi Li, Mi Ok Lee, Brian W. Davis and Andersson Leif, Texas A&M, College Station, TX

HMGA2 is a transcriptional regulating factor that does not possess any documented transcriptional activity. This gene is implicated in numerous biological roles, including growth, cancer development, craniofacial morphology in mammals, beak size in Galapagos finches. Natural variation in *HMGA2* has been associated with height in humans, as well as overall body size in numerous species such as mice, dogs, pigs, rabbits, and horses. A naturally occurring deletion in *HMGA2* is responsible for the dwarf mutation “pygmy” in mice. However, due to the intergenic location of a portion of the deletion, and proximity to other genes, this mutation may affect genes other than *HMGA2*. This obfuscation of the actual function of *HMGA2* requires the creation of a more finely targeted, induced mutation. In this study, *HMGA2* knockout mice were generated using CRISPR/Cas9 to create a stop codon in the first exon, avoiding the possibility of affecting other loci. Phenotypic analysis was executed in heterozygous and homozygous *HMGA2* knockouts, including weight, body composition, and *in vivo* micro-CT analysis utilizing computational photography to assess skeletal and craniofacial alteration. Comparisons were made between this line, and those of the natural mutation, indicating the ablation of *HMGA2* function is solely responsible for the phenotypic changes in this novel line.

P0015: Gene Editing/CRISPR

Efficient Editing PMEL17 Gene in DF1 Cells Using CRISPR/CAS9

Huijuan Bi, China agriculture university, Beijing, China

PMEL17 (premelanosome protein 17) is a pigment cell-specific integral membrane glycoprotein and its mutation resulting in hypopigmentation have been described in White Leghorn, which are covered with white plumage. In China, due to the impact of traditional culture, consumers prefer colored feather chicken. Therefore, the aim of this study is to knockout the PEML17 gene in DF1 cells and expect to obtain colored feather chickens with good production performance using the CRISPR / Cas9 technology. In the present study, four sgRNAs with varied knockout efficiencies targeting the different exons of PMEL17 were designed. Two most efficient sgRNAS named G3 and G4 were selected and the target sites were cloned into pMD19-T plasmid for sequencing. Sequencing analysis revealed that there were 9 mutational classes with 1-15bp deletions in G3 site and 67% of these deletions were not multiples of 3bp, which could induce frame shift mutations in the genome. More complex mutations were detected in the G4 site, the largest one contained 56bp deletion. Apart from the deletion mutation, 17% of them were single nucleotide mutation. Notably, although various deletion mutations were detected, no insertion mutations were found in the analyzed sequences. To test whether off-target effects occurred in these gene-edited cells, 7 potential off-target sites were analysed using T7EI assay and no detectable off-target mutations were found in these loci. These results showed that PMEL17 gene can be efficiently knockout using CRISPR/Cas9 system in DF1 cells.

Keywords: CRISPR / Cas9 technology□PMEL17 gene□genome editing

P0016: Methods: Bioinformatics

Using GDR for Translational Research

Sook Jung[1], Taein Lee[2], Chun-Huai Cheng[2], Ping Zheng[1], Jing Yu[1], Jodi L. Humann[1], Ksenija Gasic[3], Morgan Frank[1], Kristin Scott[1], Katherine M. Evans[4], Cameron Peace[1], Lisa DeVetter[1], Jim McFerson[4], Michael Coe[5], Michael Kahn[1] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Clemson University, Clemson, SC, (4)Washington State University, Wenatchee, WA, (5)Cedar Lake Research Group, Portland, OR
The Genome Database for Rosaceae (GDR, https://www.rosaceae.org) provides online resources to facilitate basic, translational and applied research for the many fruit, nut and ornamental crops belonging to the economically important Rosaceae family. Fully integrated data includes curated genome sequences, genes, genetic maps, markers, QTL, genes, transcripts, germplasm, and publications, made accessible to browse, query and download through easy-to-use web interfaces and tools. In this demo, we will highlight how to use GDR for translational research. Questions will include how to find genomic regions that are associated with a trait of interest using various data, such as genes/transcripts, QTL, marker and synteny data and how to find polymorphic markers in the genomic region of interest. The demo of this Tripal-powered database will illustrate, step-by-step, how to use various search pages and tools to find information relevant for translational research.

P0017: Methods: Bioinformatics
Genomic Data Management System (GOBII-DMS)
Yaw A. Nti-Addae[1], Star Yanxin Gao[1], Philip Glaser[1], Joshua Lamos-Sweeney[1], Kevin Palis[1], Syed Raza[1], Angel Villahoz-Baleta[1], Deb Weigand[1], Elizabeth Jones[1], Ed Buckler[2], Jean-Luc Jannink[2], Lukas Mueller[3], Kelly Robbins[4], Mark E Sorrells[4], Qi Sun[4] and Susan McCouch[4], (1)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (2)USDA-ARS, Cornell University, Ithaca, NY, (3)Boyce Thompson Institute, Ithaca, NY, (4)Cornell University, Ithaca, NY
Genomic Open-source Breeding Informatics Initiative (GOBII) is a global community of multidisciplinary teams at Cornell University, USDA, the Boyce Thompson Institute, and International Maize and Wheat Improvement Center (CIMMYT), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), and International Rice Research Institute (IRRI), James Hutton Institute (JHI) and Diversity Array Technology (DArT). Our mission is to transform breeding by developing genomic data management systems, putting the tools in place to enable the implementation of genomic and marker assisted selection as part of routine breeding programs for staple crops in the developing world. The initial main crops are: rice, wheat, maize, sorghum and chickpea in South Asia and Sub-Saharan Africa expanding to many crops worldwide.
GOBii-DMS v1 comprises of curator/data loading desktop application, web-client data extraction, restful API communicating with other applications such as KDCompute and Galaxy. Key features and API integration with other systems will be demonstrated.

P0018: Methods: Bioinformatics
Utilities of OrthoDB and BUSCO
Felipe A. Simão, University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland
Orthology is a cornerstone of comparative genomics, offering evolutionarily-qualified hypotheses on gene function by identifying “equivalent” genes in different species. The OrthoDB catalogue of orthologues represents a comprehensive resource of comparative genomics data to help researchers make the most of their newly-sequenced genomes. OrthoDB’s sets of Benchmarking Universal Single-Copy Orthologs, BUSCO, provides a rich source of data to assess the quality and completeness of these genome assemblies and their gene annotations. These resources and tools enable improved and extended orthology-based genome annotation and interpretation in a comparative genomics framework that incorporates the ever growing numbers of sequenced genomes. Such comparative approaches are well established as immensely valuable for gene discovery and characterization, helping to build resources to support biological research. The goal of this demonstration will be to introduce researchers to OrthoDB resources and tools, including the web interface for running the BUSCO genomics data quality assessment tool.

P0019: Methods: Bioinformatics
Is the Joint Variant Discovery Approach Appropriate for Calling Variants in RNA-Seq Experiments?
Jean-Simon Brouard, Université de Sherbrooke, Sherbrooke, QC, Canada
The Genome Analysis Toolkit (GATK) is a popular set of programs for discovering and genotyping variants from high-throughput-sequencing data. Versions 3.0 and above of GATK offer the possibility of calling DNA variants on cohorts of samples using the HaplotypeCaller algorithm in GVCF mode. With this approach, variants are called individually on each sample using the –ERC GVCF mode, leading to the production of one gVCF file per sample that lists genotype likelihoods and annotations for each site in the genome. In a second step, variants are called through a joint genotyping analysis from gVCF files of all samples. This strategy is more flexible and reduces computational challenges in comparison with the traditional joint discovery workflow. Although the current GATK recommendation for RNA sequencing (RNA-Seq) is to perform variant calling from individual samples, using a GVCF workflow in RNA-Seq could provide substantial advantages. That workflow has not been validated, however. In accordance with the GATK best practices for variant calling on RNA-Seq data, we compared the per-sample and the joint genotyping approaches using paired samples from 56 cows genotyped with RNA-Seq data derived from whole primary macrophage transcriptomes, genotyping-by-sequencing (GBS) data, and Bovine SNP50 BeadChip data. Our results indicate that the per-sample and the joint genotyping approaches perform similarly in terms of sensitivity (>90%) and precision (>70%). Our results also indicate that RNA-Seq genotypes with high accuracy (>98%) can be obtained with RNA-Seq data. In addition, we found that a sizeable proportion of discrepancies between the GBS variant calls and the RNA-Seq variant calls would be explained best by RNA-Seq editing events. This study suggests that joint genotyping is a suitable variant-calling method when conducting RNA-Seq experiments.

P0020: Methods: Bioinformatics
CitrusID: A Web-Based Software to Identify Citrus Variety
Lun Wang, Huazhong agricultrual University, Wuhan, China
CitrusID is comprehensive specific genomic admixture pattern database of the citrus family including 50 mandarins, 287 pummel, 11 citron, 12 ichang papeda, 16 Atlantia, 12 poncirus, 9 kumquat and hybrids such as 34 sweet orange, 25 sour orange, 7 grapefruit, 10 lemons, as well as 16

others unknown-origin hybrids .We develop a web-based algorithm to characterize the genetic characteristics and identify the cultivar/variety. It aimed to confirm the tested sample category it belongs or the type of citrus species and show the genetic components by providing the genetic contributions of each basic species. Database URL: (http://211.69.140.136/orange/)

P0021: Methods: Bioinformatics

The Role of Public Sequence Repositories in Teaching Bioinformatics

Pushpinder Singh Bawa, IBAB, India, **Dr. Shubham Dixit**, Institute of Bioinformatics and applied biotechnology, Bangalore, India, Samathmika Ravi, Institute of Bioinformatics and Applied Biotechnology, Bengaluru, Karnataka, India and Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Impressive advances in sequencing technologies has made bioinformatics an indispensable discipline under biology. Data generation, which was a huge bottleneck only a decade ago, is now proliferating public repositories with petabytes of data from diverse NGS applications. Today, biological research can begin without ever entering a laboratory. The volume and velocity with which data are pouring into the public repositories is leaving no room for the research communities to equip themselves with skills or computing infrastructure that is needed to make sense of the data. There is a dire need for trained bioinformatician in biology departments across the globe. Yet, there is lack of trainers and facilities to meet this demand. Virtual classrooms, public repositories and cloud computing infrastructure can be harnessed to take full advantage to scale training programs in bioinformatics. Since 2010, IBAB has extensively used public sequence repositories to provide hands on training in bioinformatics using in-house computing infrastructure. Here, we will be presenting our efforts towards use of cloud computing infrastructure to expand our hands on training programs to remote classrooms with no computing infrastructure of their own. The future of biology will, for the most part, be driven by mining data stored somewhere in the cloud by someone in the distal part of the world.

P0022: Methods: Bioinformatics

Developing a Plant Genome Database PGDBj in the NGS Era

Jeffrey A. Fawcett[1], Hideki Hirakawa[1], Mitsuyo Kohara[1], Andrea Ghelfi[1], Tsunakazu Fujishiro[1], Hisako Ichihara[2], Akihiro Nakaya[2], Sachiko Isobe[1] and Satoshi Tabata[1], (1)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (2)Graduate School of Medicine, Osaka University, Suita, Osaka, Japan

With the increasing amount of genome sequence data being generated, databases and webtools have important roles in facilitating various analyses using the genomic data and disseminating knowledge gained by such analyses. At Kazusa DNA Research Institute, we have been sequencing the genomes of a wide range of plant species. The challenge we are facing now is to provide the communities of each species with the data and tools that will allow them to analyze polymorphic data (e.g. SNPs) and to identify loci associated with various important traits. Here, we will present and discuss our efforts in developing a new version of a plant genome database, PGDBj, which was previously developed as more of a portal website. This new version will integrate data such as the genomic sequences, gene annotations, homology information, polymorphic data, DNA markers, and QTLs of various plant species. Emphasis will be on species sequenced by Kazusa and their closely related species, starting with those such as *Lotus japonicus*, strawberry, sweet cherry, and subclover. Our goal is to provide detailed information and dedicated services to the communities of certain specific species, while at the same time providing a platform for plant genomic studies to the broader plant research community.

P0023: Methods: Bioinformatics

Methods for Resolving Bird Genomes using Linked-Read Sequencing

Remi-Andre Olsen, Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna, Sweden

The SciLifeLab National Genomics Infrastructure (NGI) is a national platform providing DNA sequencing services to Swedish researchers in fields such as human health and bio-diversity. With regard to *de novo* sequencing of animal genomes we are in a unique position to evaluate emerging technologies and to develop methods for laboratory automation and bioinformatics pipelines that would benefit the wider genomics community.

We present the sequencing and assembly of 11 samples of bird-of-paradise species using linked-reads from the 10X Genomics Chromium system, assembled with Supernova. The aim of this work is to provide long-range information needed to identify female W-linked sequences and other highly repetitive regions. We have developed a new visualization method for Supernova assembly results via a module for the "MultiQC" reporting tool. We show how our bioinformatics pipeline (NGI-NeutronStar) implemented in Nextflow automates multiple Supernova runs and subsequent assembly evaluation programs.

We were able to produce highly contiguous genome assemblies of several birds-of-paradise. We show how input parameters such as mean molecule lengths and sequencing depth in Supernova can affect the quality of resulting assemblies. We show that linked-read technologies are a promising method for *de novo* sequencing of bird genomes.

P0024: Methods: Bioinformatics

Expert Curation of Plant Proteins in UniProtKB/Swiss-Prot

Damien Lieberherr[1], Sylvain Poux[1] and UniProt Consortium, (1)SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

The UniProt KnowledgeBase (UniProtKB) provides a single, centralized, freely available resource for protein sequences and functional information. For Arabidopsis, our main targets for expert annotation are proteins with some functional characterization and most of them are now included in UniProtKB/Swiss-Prot. Expert curation combines the manually verified sequence with experimental evidence derived from biochemical and genetic analyses, 3D-structures, mutagenesis experiments, information about protein interactions and post-translational modifications. Besides harvesting, interpreting, standardizing and integrating data from literature and numerous resources, curators are also checking, and often correcting, gene model predictions.

Our annotation program has been actively collaborating with other resources, such as Araport, the Arabidopsis portal. We provide Araport with all the gene model corrections that we introduced on the bases of our trans-species family annotation. We are also completing the knowledgebase by importing missing information from EnsemblPlants.
The UniProt consortium is also actively involved in GO annotation and 19,000 manual annotations has been added to more than 4700 plant proteins. Experimental peptides from high-throughput proteomics experiments that uniquely match the product of a single gene are used to generate annotations describing post-translational modifications and protein processing events. UniProtKB serves as a central hub for biomolecular information with access to more than 100 other resources, such as nucleotide sequence database, 3D protein structure databases, InterPro or MODs.

P0025: Methods: Bioinformatics

Comparison of Bioinformatics Software Tools using RNA-Seq Data from Three Different Populations of Chickens with Wooden Breast Disease

Jose Daniel Chazi Capelo and Behnam Abasht, University of Delaware, Newark, DE

Currently, there exist several bioinformatics tools for processing sequence reads generated from RNA-seq studies. Although the choice of a program in a RNA-seq study is dependent on the researcher's preference, information about comparison of the suites is limited. Moreover, studies on validation of results using non-sequencing-based techniques are lacking. This study aimed at comparing different bioinformatics programs using 22 RNA-seq samples comprising affected and unaffected groups from three distinct broiler populations with Wooden Breast Disease. Two aligners (HISAT, STAR) and two differential expression (DE) tools (Cuffdiff, DESeq) were used to generate four combinations of pipelines for the 3 datasets. Each combination generated a list of significant genes for each population. All the lists were evaluated for overlapping genes among the three populations to determine behavior and concordance. Although both aligners produced satisfactory read coverage; STAR had a higher percentage of mapping-rate (88%) than HISAT (76%). Combinations using Cuffdiff produced higher percentages of overlapping genes (25%) than those with DESeq (21%). Almost 100% of overlapping genes showed concordant expression patterns across all the three datasets for all combinations. Results also exhibited DE tools are not susceptible to aligners and produced similar outcomes. However, DE tools that used the same aligner produced different outcomes. Cuffdiff presented susceptibility to number of replicates and had better performance in larger groups. To corroborate RNA-seq study, an independent dataset was generated using a non-sequencing-based (Nanostring) technology; yielding a correlation of 0.85 between the two technologies. Overlap significant genes between Nanostring and RNA-seq also confirmed the behavior.

P0026: Methods: Bioinformatics

Prometheus: Omics Portals for Interkingdom Comparative Genomic Analyses

Gunwhan Ko[1], Insu Jang[2], Hyeongmin Kim[3], Young Mi Sim[1] and **Yong-Min Kim**[4], (1)Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, South Korea, (2)Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea, Republic of (South), (3)Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Deajeon, South Korea, (4)Korea Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejon, Korea, Republic of (South)

Functional analyses of genes are crucial for unveiling biological responses, for genetic engineering, and for developing new medicines. However, functional analyses have largely been restricted to model organisms, representing a major hurdle for functional studies and industrial applications. To resolve this, comparative genome analyses can be used to provide clues to gene functions as well as their evolutionary history. To this end, we present Prometheus (http://prometheus.kobic.re.kr), web-based omics portal that contains more than 17,215 sequences from prokaryotic and eukaryotic genomes. This portal supports interkingdom comparative analyses via a domain architecture-based gene identification system, Gene Search, and users can easily and rapidly identify single or entire gene sets in specific pathways. Bioinformatics tools for further analyses are provided in Prometheus or through BioExpress, a cloud-based bioinformatics analysis platform. Prometheus suggests a new paradigm for comparative analyses with large amounts of genomic information.

P0027: Methods: Bioinformatics

A Bayesian Method for Prediction of Hybrid Performance Based on Principal Components Analysis

Haixiao Dong and Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

Genomic selection (GS) has been widely used to accelerate breeding with the development of high-throughput genotyping and phenotyping technologies. Several GS methods have been developed, including genomic BLUP (Best Linear Unbiased Prediction), ridge regression, Bayesian methods. The Bayesian methods have the advantage of taking different prior distributions and were demonstrated to be superior to other methods in some circumstances. The challenge is computing speed, especially for prediction of hybrid performances because the factors not only include additive effects, but also non-additive effects such as dominance. Here we proposed a model to predict hybrid performances in framework of Bayesian method based on principal component analysis (PCA). Genotypes were coded into additive and dominant genotypes and then conducted with PCA separately. The additive Principal Components (PC)s and the dominant PCs were fitted as random effects in a Bayesian model to predict hybrid performances. The computing time was reduced in the order of magnitude. Stochastic validation demonstrated that the PCA based methods had very similar prediction accuracy as the Bayesian method that takes the original additive and dominant genotypes as the random variables.

P0028: Methods: Bioinformatics

Genome Archive(R): Standardized Genome Repository for Supporting Large-Scale Genome Analyses

Jongsun Park and Hong Xi, InfoBoss Co., Ltd., Seoul, South Korea

Owing to Next Generation Sequencing (NGS) technologies. More than 15,000 whole genome sequences have been sequenced till now ranging from bacteria to human. There are several major data archives, such as NCBI, DDBJ, and EBI, for collecting them. These databases serve structured file of genome information, such as fasta and gff3 files, web interface to access data, as well as analysis tools, including BLAST.

There are another databases which archive genome sequences with their own pipelines, for example, Ensembl covering many Eukaryotic species and Phytozome archiving plant genomes usually sequenced by Joint Genomic Institute. These databases provide better bioinformatics tools to dissect them efficiently. However, all these databases do not contain all available genomes, requiring additional efforts to analyze the genomes from different repositories. Especially, gff3 files from different databases have slightly different formats, which is another hurdle to integrate them for researches. To overcome this problem, we developed a standardized genome database, named as GenomeArchive® (http://www.genomearchive.info). Currently it contains six genome databases: Plant Genome Database, Fish Genome Database, Insect Genome Database, Fungal Genome Database, Nematode Genome Database, and Bacterial Genome Database. Due to a standardized form of genome sequences in Genome Archive®, these genome databases can share genome data freely through GlobalScrap®, which is a simple on-line cart to collect and to analyze sequences. Moreover, GenomeArchive® have been utilized for customized genome databases including Psedostellaria Database (http://www.pseudostellaria.net/), Salix Genome Database (http://www.salixgenome.info/), and Coffee Genome Database (http://www.coffeegenome.info/). GenomeArchive® can be a generated platform not only for supporting diverse bioinformatics analyses but also for presenting genome data efficiently through web interfaces.

P0029: Methods: Bioinformatics

A Robust and Efficient Method for the Phenotypic Analysis of Plant Breeding Experiments

Martin P. Boer[1], María Xosé Rodríguez-Álvarez[2], Paul H.C. Eilers[3], Julio Velazco[4], Marcos Malosetti[1] and Fred A. van Eeuwijk[4], (1)Wageningen University & Research - Biometris, Wageningen, Netherlands, (2)BCAM, Bilbao, Spain, (3)Erasmus University Medical Centre, Rotterdam, Netherlands, (4)Wageningen UR - Biometris, Wageningen, Netherlands

Adjustment for spatial trends in plant breeding field trials is essential for efficient evaluation and selection of genotypes. Current mixed model methods of spatial analysis of field trials are based on a multi-step modelling process where global and local trends are fitted after trying many candidate spatial models. This process of model selection is often quite complicated and time consuming, since for each trait and each environment different optimal models need to be selected.

A recently developed method, referred to as SpATS, uses two-dimensional P-splines to model the spatial trends in the breeding experiments. The improvements in precision and the prediction of genotypic values produced by the SpATS model are equivalent to those obtained using the best fitting standard spatial models for each trial.

One of the key advantages of the SpATS method is that no model selection is needed, and therefore easily runs over multiple traits, multiple environments, or multiple time points for high throughput phenotyping experiments. In addition, the SpATS method is very robust, which makes it possible to fully automate the phenotypic analyses. An R-package, called SpATS, implementing the method is available on CRAN.

P0030: Methods: Bioinformatics

Haplotyping of Full-Length Transcript Reads from Long-Read Sequencing Can Reveal Allelic Imbalances in Isoform Expression

Elizabeth Tseng[1], Timothy P.L. Smith[2], Sarah B Kingan[1], Stefan Hiendleder[3], Ruijie Liu[3] and John Williams[3], (1)Pacific Biosciences, Menlo Park, CA, (2)USDA, ARS, USMARC, Clay Center, NE, (3)Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia

Full-length transcript sequencing using long read technology has enabled researchers to characterize alternative splicing events and improve genome annotation. The Pacific Biosciences Iso-Seq method, which can produce high-quality isoform sequences of 10 kb and longer, has been used to annotate many important plant and animal genomes. Here, we develop an algorithm called IsoPhase that post-processes Iso-Seq data to retrieve allele-specific isoform information. IsoPhase first piles up the full-length circular consensus sequences (CCS) of all the isoforms of a gene and calls SNPs using a modified version of PacBio's minor variant calling algorithm, Juliet. It then infers haplotypes based on the phasing information the CCS sequences provided. The output defines the inferred haplotypes for each transcript and estimates the relative abundance of each allele.

We tested IsoPhase on simulated data to determine false discovery rate (FDR) at various sequence read characteristics, genome ploidy, and coverage. For both diploid and tetraploid genomes, we show that error rates, consistent with CCS error rates, result in very low FDR.

Finally, we describe the application of IsoPhase to a haplotype-resolved genome assembly and multiple fetal tissue Iso-Seq dataset from a F1 cross of Angus x Brahman cattle subspecies. IsoPhase-called haplotypes were validated by the phased assembly and demonstrate the potential for revealing allelic imbalances in isoform expression.

P0031: Methods: Bioinformatics

JBrowse and AWS S3: Cheap and Easy Genome Browsing

Scott Cain, Ontario Institute for Cancer Research, Medina, OH and the GMOD Consortium

JBrowse is widely used GMOD (Generic Model Organism Database, http://gmod.org/wiki/JBrowse) software for displaying a variety of genomic feature data types. Here we present a method for implementing JBrowse in a very low maintenance fashion by loading the data and software for JBrowse into Amazon Web Service's S3 data storage service and serving the web pages and data directly out of S3, without the need for any other web server. We outline the methods for implementation as well as the issues that implementers may want to consider, including cost and security.

P0032: Methods: Bioinformatics

Evolutionary Computation to Improve Genomic Selection Accuracy

Joseph O'Neill and **Travis W. Banks**, Vineland Research and Innovation Centre, Vineland Station, ON, Canada

For complex traits where many loci have small effects on the phenotype, genomic selection (GS) for plant improvement is often used, rather than individual QTL-associated markers. With GS, phenotypic or breeding values are estimated using multiple genetic markers from across the genome, such that all QTL are in LD with at least one genetic marker. Using real and simulated data sets, we show phenotypic prediction can be improved by using a particular subset of genetic markers, instead of all markers from genome-wide microarray panels. We develop a metric

from encoding genotype calls numerically and performing a PCA in order to evaluate the correlation between the first PCA values and the phenotype. This metric serves as a fitness function in an evolutionary computation-based search for the ideal marker subset which is then used to build our predictive model. Application of this approach to simulated genotyping-by-sequencing data is also presented.

P0033: Methods: Bioinformatics

New Automatic Gene Finding Tool for Compact Eukaryotic Genomes - GeneMark-EX.

Alexandre Lomsadze, Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering,Georgia Institute of Technology, Atlanta, GA, Tomas Bruna, School of Biology, Georgia Institute of Technology, Atlanta, GA and Mark Borodovsky, Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA

Current tools for genome wide prediction and annotation of protein coding genes are expected to be fully automatic. Developers of genome annotation pipelines are seeking for optimal ways of integration of spliced alignments of transcripts (RNA-Seq) and mapped to genome proteins into the training and prediction steps of *ab initio* gene finders. All parameters of such comprehensive multi-layer tools, including the species specific parameters of statistical models are expected to be generated automatically. Performance of the annotation pipelines still critically depends on the volume and reliability of the input data (assembled genome, RNA-Seq reads, protein database).

We present a fully automatic integrated tool, GeneMark-EX, for genome annotation that shows robust performance across the input data of various size, structure and quality. The algorithm selects the approach to parameter estimation depending on the volume, quality and features of the input data, size of RNA-seq dataset, phylogenetic position of the species, degree of assembly fragmentation. It is able to automatically modify the HMM architecture to fit the features of the genome in question and to integrate transcript and protein information into the process of gene prediction.

This algorithm was designed for the species that have compact genomes (up to 400 Mb of meaningful sequence after repeat masking) and infrequent alternative splicing. For the species with high frequency of alternative splicing, this tool would provide a training set for generating parameters of the AUGUSTUS gene prediction algorithm; this logic of integration of the two efficient tools has already been implemented in the BRAKER pipeline.

P0034: Methods: Bioinformatics

Fish Genome Database: A Standardized Fish Genome Repository for 153 Fish Species

Jongsun Park and Hong Xi, InfoBoss Co., Ltd., Seoul, South Korea

Owing to next generation sequencing technologies as well as research interests in genome sequences, more than 100 fish species are available in public. However, there is no integrated repository for fish genomes. For example, Ensembl (http://www.ensembl.org/; Release 90) contains around 13 fish genomes, which is smaller number of genomes archived in NCBI genome database. 66 teleost fish genomes sequenced for understanding evolution of immune systems are not included in NCBI as a genome. Moreover, some genomes published in Giga Science journal are not deposited in any databases. It requires additional efforts to collect fish genomes from different sources for researches. To reduce these additional efforts and to monitor current status of fish genomes, we constructed Fish Genome Database (http://www.fishgenome.info/) based on GenomeArchive®. Currently, Fish Genome Database (Release 1) contains 155 fish genomes from 153 species covering 15 out of 57 orders. With the aid of GlobalScrap®, it also provides several bioinformatics analysis tools which can be executed on the web with any sequences deposited in Fish Genome Database. Due to 66 fish genomes sequenced at low coverages, 50 out of 155 fish genomes have gene model. Total number of ORFs from them is 1,888,497, of which average number per genome (37,769.94) is smaller than that of plant genomes (42,771.14). Based on InterProScan program, more than 80% of fish gene were annotated with functional domains. Moreover, GATA transaction factors and cytochrome P450s from 50 fish genomes were identified and presented in the database. With keeping update of fish genome sequences, Fish Genome Database will be a useful repository for comparative fish genomic researches.

P0035: Methods: Bioinformatics

Pathway Enrichment for Model and Non-Model Organisms with Pathway Inspector

Luca Bianco[1], Samantha Riccadonna[1], Sook Jung[2], Dorrie Main[2] and Paolo Fontana[1], (1)Research and Innovation Centre, Edmund Mach Foundation, San Michele all'Adige, Italy, (2)Washington State University, Pullman, WA

Functional analysis of high-throughput data is a fundamental step in systems biology investigations. To fully exploit the information encoded in complex biological systems, data analysis tools should be able to combine omics experiments and background knowledge to describe the interactions among proteins and small molecules and to improve the limited knowledge available on gene functions.

Pathway Inspector [1] is a web application helping researchers to find patterns of expression in complex RNAseq experiments. The tool combines two standard approaches for RNAseq analysis: the identification of differentially expressed genes and a topology-based analysis of enriched pathways. Pathway Inspector is provided with ad-hoc interactive graphical interfaces simplifying the discovery of modulated pathways and the integration of the differentially expressed genes in the corresponding pathway topology. Pathway Inspector is now integrated into the Genome Database for Rosaceae (https://www.rosaceae.org/) [2] and can perform its analysis based on the GDRCyc pathways. Pathway Inspector is freely available at https://pathwayinspector.fmach.it/ and https://www.rosaceae.org/tools/pathway_inspector

References:

1. L.Bianco et al., Pathway Inspector: a pathway based web application for RNAseq analysis of model and non-model organisms, Bioinformatics, Volume 33, Issue 3, 1 February 2017, Pages 453–455, https://doi.org/10.1093/bioinformatics/btw636
2. S. Jung, et al., The Genome Database for Rosaceae (GDR): year 10 update. Nucl. Acids Res. (2014) Jan 1; 42(1):D1237-44

P0036: Methods: Bioinformatics

Stilbenoid Prenylation Pathway Discovery in Peanut using Targeted Transcriptomics

Sheri Sanders[1], Ram Podicheti[2], Tianhong Yang[3,4], Lingling Fang[4], Srinivas Jayanthi[5], Gayathri Rajan[2], Thallapuranam Krishnaswamy Suresh Kumar[5], Fabricio Medina-Bolivar[6] and Keithanne Mockaitis[1,7], (1)National Center for Genome Analysis

Support, Pervasive Technology Institute, Bloomington, IN, (2)School of Informatics and Computing, Bloomington, IN, (3)Arkansas Biosciences Institute and Department of Biological Sciences, Jonesboro, AR, (4)Arkansas State University, State University, AR, (5)Department of Chemistry and Biochemistry, Fayetteville, AR, (6)Arkansas Biosciences Institute, Jonesboro, AR, (7)Department of Biology, Indiana University, Bloomington, IN

In peanut, *Arachis hypogaea*, defense responses to biotic and abiotic stresses include the synthesis of prenylated stilbenoids, with over 45 such compounds identified to date. The diversification of secondary metabolites in plants is expanded by prenylation activities, and in recent studies this modification has been shown to enhance biological activities of polyphenolic compounds.

We describe our discovery of genes responsible for stilbenoid prenylation* as well as studies underway to understand the regulation of these metabolic programs in peanut. Sequencing RNA from a well-characterized peanut hairy root system, we built a transcriptome reference and correlated transcripts with metabolites produced over a time course of elicitation. Transcripts encoding candidate enzymes were identified and characterized functionally by heterologous transient expression. Prenyltransferases we call AhR4DT-1 and AhR3'DT-1 catalyze distinct reactions, and our studies suggest that these act in the first committed steps that convert non-prenylated into prenylated stilbenoids.

Here we highlight the functional transcriptomics that led to these discoveries, and our ongoing approaches to find other genes that act in the regulation of this defensive metabolic program. Our identification of the first plant stilbenoid-specific prenyltransferases advances the understanding of this specialized gene family, and contributes some of the functional definition that is needed generally to refine the annotations of plant genomes.

*Yang T, Fang L, Sanders S, Jayanthi S, Rajan G, Podicheti R, Kumar TKS, Mockaitis K, Medina-Bolivar F, 2017. *JBC, in press.* Stilbenoid prenyltransferases define key steps in the diversification of peanut phytoalexins.

P0037: Methods: Bioinformatics

Maximizing Gene Prediction using Traditional Gene Finding Pipelines

Arun S. Seetharam and Eve Syrkin Wurtele, Iowa State University, Ames, IA

Many genomes are annotated by combining direct RNA-seq alignments and conserved orthologous genes as evidence for *ab-initio* machine-learning. Here, we systematically evaluate the ability of current *ab initio* software, Maker and BRAKER to predict genes with various characteristics (GC content, amino acid usage, length, exon number, and others). Using defined model species and benchmark datasets, we predicted as few as ten percent of some transcripts of particular characteristics by some commonly-used ab initio approaches. We noticed significant differences in the sensitivity (Sn) and specificity (Sp) scores between the classes of gene. The performance of *ab-initio* gene predictions was highly dependent on the data source. In this study, we optimize *ab-initio* gene prediction for various gene features, thus facilitating the identification of a wider variety of genes.

P0038: Methods: Bioinformatics

Solving the Challenges of Complex Genome Analysis Collaborations on-line using XSEDE Resources

Carrie Ganote[1], Sheri A. Sanders[2], Le-Shin Wu[3], Tom Doak[1] and Keithanne Mockaitis[2], (1)Indiana University, Bloomington, IN, (2)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (3)NCGAS, Indiana University, Bloomington, IN

Whole genome reference and functional genomics projects most often benefit from a diversity of expertise and the integrated contributions of multiple research groups. The importance of web-enabled data sharing and open access software to progress in genome research is indisputable. Dissemination of genome resource information through internet-based resources, especially customizing and encouraging the optimal use of analysis tools to serve a specific research community, often lags behind data generation. This lag can inhibit biological interpretations and downstream experimentation, much of which should be undertaken before a genome resource project is completed. File systems for storing data and analysis outputs of today's project standards must be large and secure, and must offer sustained access to fulfill the hosting requirements an active and dispersed research community needs. Users must interact with a computing resource powerful enough to get their jobs done, and sites must be expandable and flexible to accommodate a growing demand for intra-genera comparisons and pan-genomics of a species. We have been using NSF XSEDE computational resources including Jetstream along with the High Performance Computing systems of Indiana University to meet these challenges for a variety of collaborative plant genomics studies. Currently these efforts are impacting metabolic gene discovery in the tetraploid *Arachis hypogaea* and whole genomic reference studies of the tetraploid *Coffea arabica* and its diploid progenitors. Here we show practical examples of XSEDE resource use and development that may benefit other genomic research groups seeking to increase the effectiveness of their computing and collaboration.

P0039: Methods: Bioinformatics

Apollo a Tool for Visually Editing Genome Annotations

Nathan A Dunn[1], Deepak R. Unni[2], Monica C. Munoz-Torres[1], Eric Yao[3], Paul J Hale[4], Ian Holmes[5], Christine G. Elsik[2] and Suzanna Lewis[1], (1)Lawrence Berkeley National Laboratory, Berkeley, CA, (2)Division of Animal Sciences, University of Missouri, Columbia, MO, (3)Department of Bioengineering, Berkeley, CA, Berkeley, CA, (4)Mouse Genome Informatics, Bar Harbor, ME, (5)Department of Bioengineering, Berkeley, CA

Manually curation of genome annotations using experimental data and existing predictions from similar genomes is an essential part of the genome annotation process. To alleviate the burden of a manual process, we provide the web-based Apollo (https://github.com/GMOD/Apollo) genome annotation editor, which allows users to collaboratively revise and edit the structure and function of predicted genomes manually, graphically, and in real-time (similar to Google Docs).

Apollo is built on top of the JBrowse genome browser and its plugin ecosystem. Though Apollo can be installed and run locally (and via Docker https://github.com/GMOD/docker-apollo) it also scales to support hundreds of users and organisms. Access to genomes is controlled with fine-grained permissions (e.g. administrator, curator, public). To aid in editing, Apollo offers its own collapsible side-'Annotator Panel', that provides an alternate view of annotations, sequences, and tracks as well as numerous screens for administration and annotation reports. Custom tabs may also be readily configured. Annotations can be exported within the interface in FASTA, GFF3, or Chado. Additionally,

annotations retain a visual revertible history of structural edits including the ability to change the type of genome feature. To support integration into existing workflows, we expose the existing suite of web services that drive user-interface functionality. These web-services have been leveraged to integrate with Docker and the Galaxy platform. Additionally, custom tabs can be added via simple configuration.
We continue to support workflows by adding the ability to add organisms and track data remotely, including indexed FASTA genomes. To support a couple of separate initiatives (Alliance of Genome Resources and the Monarch Initiative), we've expanded our track services to deliver JSON snippets that represent small regions of tracks or specific genome elements that can also be cached. This has allowed the ability to easily add embeddable and fast "overview" SVG widgets to genome feature pages (https://github.com/GMOD/GenomeFeatureComponent).
Increasing Apollo's ability to visually explore and analyze genome data, two major undertakings are currently under development. The first is the ability to project genomic feature regions, removing intra- and intergenic regions to provide a more information-rich visualization of the genome. This will allow, for example, hiding introns or showing only regions with relevant genes. While projection is underway, we are also providing an npm projection library to be integrated within JBrowse (https://github.com/gmod/linear-projection) to directly support this functionality. Second, is the ability to visualize variant data and to annotate their predicted effects, primarily on coding regions. New technology trends and scientific paradigms point to new needs in genomic analytic tools to leverage information about variants that impact human health. Driven by a growing need to identify disease causing variants across diverse groups, we are working towards providing full functionality in genomic variant analysis and curation.
Apollo is used in over one hundred genome annotation projects around the world, ranging from annotation of a single species to lineage-specific efforts supporting the annotation of dozens of species at a time.
Project page: http://genomearchitect.org/

P0040: Methods: Bioinformatics
Inflated False Positives behind a Beautiful QQ Plot in GWAS
Meijing Liang[1,2] and Zhiwu Zhang[1,2], (1)College of Animal Sciences, Northeast Agricultural University, Harbin, China, (2)Dept. of Crop and Soil Science, Washington State University, Pullman, WA
False positives are misleading in discovery transformation and create more false positives in research downstream. False positive rates are usually controlled by statistical testing threshold (Type I error). However, non-replicable discoveries in Genome-Wide Association Studies (GWAS) were beyond the scope of Type I error. Several additional causes were diagnosed, including the chain of one error leading to more errors. In this study, we identified another cause due to QQ plot inflation over expectation. Both real data and simulated data were used for the demonstration. We reanalyzed the real data of lung cancer in asian populations with General Linear Model (GLM) and Mixed Linear Model (MLM). Both models used the first three Principal Components (PCs) derived from all the available markers to control population structure. Additionally, MLM incorporated kinship among individuals for phenotype-marker association tests. Most of the markers lined up nicely with null expectation on QQ plots by both methods. There were three markers that exceeded the expectation by using MLM and 35 markers by using GLM, suggesting GLM had more "power" than MLM. Then we selected the top 500 associated SNPs on the first twelve chromosomes identified by GLM as the Quantitative Trait Nucleotides (QTNs) to simulated an arbitrary trait with heritability of 75%. There were no QTNs from the rest chromosomes. GWAS with GLM identified 30 false associations on the second half of chromosomes, compared with one claimed by MLM. Our results suggested that a beautiful QQ plot should be interpreted with caution and MLM is recommended over GLM to control false positives in GWAS.
Keywords: False positives, Q-Q plots, GWAS

P0041: Methods: Bioinformatics
Repbase: A Comprehensive Database of Eukaryotic Repeat Sequences and Transposons
Kenji K. Kojima[1,2], Weidong Bao[1] and Oleksiy Kohany[1], (1)Genetic Information Research Institute, Mountain View, CA, (2)National Cheng Kung University, Tainan, Taiwan
Repbase is a comprehensive collection of representative repeat sequences in eukaryotic genomes. Since its first development as a database of human repetitive sequences in 1992, Repbase has been serving as a well-curated reference database fundamental for almost all eukaryotic genome sequence analyses. One main usage of Repbase is to mask repetitive sequences from the genome of interest with software such as Censor or RepeatMasker. Because most of the repetitive sequences originated from various transposons as well as integrated viruses, Repbase is also a fundamental source for the studies of transposons and their impact on the genome evolution.
One distinguished feature of Repbase is the detailed characterization/annotation of the repeat sequences that includes the classification into a particular transposon superfamily, the average identity/similarity of sequence copies to the consensus, and the reconstructed protein sequences that were originally encoded by their ancestral transposons.
Recently, we have taken large efforts for comprehensive characterization/annotation of repeat sequences from three organisms, the human *Homo sapiens*, the rice *Oryza sativa* and the zebrafish *Danio rerio*. For humans, we carefully annotated and classified ancestral repeats which are shared by many mammals or other vertebrates. Both rice and zebrafish have more than 2000 families of transposons in Repbase, although their genome sizes are much different (rice: 390Mb, zebrafish: 1.5Gb).

P0042: Methods: Bioinformatics
Hayai-Annotation: Ultra-Fast Plant Annotation Workflow
Andrea Ghelfi[1], Hideki Hirakawa[2], Kenta Shirasawa[3] and Sachiko Isobe[1], (1)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (2)Tohoku University, Kisarazu, Japan, (3)Kazusa DNA Research Institute, Chiba, Japan
Increase productivity and efficiency use of nutrients, pathogens and disease resistance, drought and heat tolerance, are essential targets for genetic improvement in plants. In order to achieve this it is critical, for molecular breeders, have a broad and accurate understanding of theirs plant protein profiles. Since complete genome sequences are becoming faster and cheaper, an automated, fast, and accurate annotation workflow should be used. Hayai-Annotation proposes to provide this service in two different interfaces. Users can choose between a WebInterface or a GUI R-package Interface. In the first case the system is provided by the Internet, the fasta file can be uploaded and the results are shown in html format, with a search bottom for users interactions and an download option. The second format is an R-package with

a Graphical Interface using RGtk2 library. Hayai-Annotation uses the free version of usearch (v10.0.240_i86linux32) against Viridiplantae taxonomy of UniProt Knowledgebase (UniprotKB). The workflow generates a table with gene name, GO enrichment, and Evidence code for each query entry identified in the database. Regarding accuracy and speed, Hayai-Annotation (R-package version) identified and properly annotated 98.4% of the sequences in 14.9 minutes (6Gb RAM, i5-2450M CPU@2.50GHzx4) using swissprot database, Viridiplantae taxonomy, as a sample test. Hayai-Annotation can provide a simple, fast and accurate annotation for plants.

P0043: Methods: Bioinformatics

Multivariate *F*-Test in Genome-Wide Association Mapping

Akio Onogi, Department of Agricultural and Environmental Biology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Bunkyo-Ku, Tokyo, Japan; JST PRESTO, Tokyo, Japan

Genome-wide association mapping (GWA) is now a widely used technique to identify quantitative trait loci (QTL) in various plant and animal species. However, GWA is inefficient for a small sample size, which is often the case in plant studies. A possible way to enhance the detection power of GWA is using multivariate information. In this study, we formulated the *F*-test for multivariate mixed effect models and evaluated it using simulation. The purpose of the multivariate *F*-test is to identify QTLs that are common to multiple variates and are undetectable by the univariate test for each variate. The advantage of the multivariate *F*-test over the univariate test varied based on three parameters: phenotypic correlation between variates (r), relative size of QTL effects between variates (a), and missing proportion of phenotypic records (m_{prop}). When m_{prop} was small, the multivariate *F*-test outperformed the univariate test as r and a differed, and as m_{prop} increased or reached a maximum, the multivariate *F*-test outperformed the univariate test as a increased. The maximum value of m_{prop} indicates a situation wherein each sample has a phenotypic record for only a single variate. Such a case would include meta-analysis or cases treating different sample groups, e.g., sex or geographical populations, as imaginary multiple variates. Our results suggest that the multivariate *F*-test is a potentially useful method to apply in GWA when multivariate data are available.

P0044: Methods: Bioinformatics

Comparative Genome Analysis of Soybean Germplasms by Optical Mapping

Claire Yik-Lok Chung[1], Alden King-Yung Leung[2], Hon-Ming Lam[1] and **TingFung Chan**[2], (1)Center for Soybean Research of the Partner SKL of Agrobiotechnology, and School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, (2)The Chinese University of Hong Kong, Shatin, Hong Kong

Optical mapping is an imaging technique for capturing enzymatic patterns along DNA molecules of hundred-kilobase scale. It has been applied to different genomic applications including genome assembly, microbial strain typing, and structural variation detection. Recently, high-throughput technology for optical mapping data generation raises interest on its applications in comparative genomics. Previous studies using optical mapping in comparative genomics mainly focuses on detecting structural variations between two genome assemblies, or between multiple genome assemblies together with a reference. However, result interpretation in the former becomes difficult when three or more genomes are analyzed. While the latter is able to analyze more genomes, it is limited by the quality and completeness of the reference. In addition, both approaches cannot handle well multiple types of variations within the same region, or regions frequently confounded by repetitive elements, inversions or other large-scale rearrangements. Such regions are usually variation hotspots that carry important biological meanings.

We introduce an analysis framework using a reference-free approach based on multiple alignment of optical map contigs. Segments of DNA molecules with signal patterns conserved among samples are grouped as collinear blocks. Samples with higher similarity as sharing more conserved regions are clustered together, providing hints for phylogeny. Blocks contributing to most variations between sample clusters are found by statistical analysis. Further annotation reveals the identities of these blocks as the most representative features. We anticipate this method can facilitate comparative genome analysis on soybean, which has a high degree of diversity between different germplasms.

(This work is supported by the UGC Area of Excellence Scheme AoE/M-403/16, the Lo Kwee-Seong Biomedical Research Fund, and the Lee Hysan Foundation.)

P0045: Methods: Bioinformatics

Mainlab Chado Module, a Tripal Module for Loading Map, Marker, QTL, Genotype, Phenotype and Germplasm Data

Sook Jung[1], Taein Lee[2], Chun-Huai Cheng[2], Stephen P. Ficklin[1], Jing Yu[1], Jodi L. Humann[1] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA

The Tripal Mainlab Chado Loader (MCL) provides data collection templates in Excel with defined metadata and data loaders with front end forms. Supported data types include organism, marker, QTL, Mendelian Trait Loci, germplasm, map, project, phenotype, genotype and their associated metadata. These data collection templates and PHP loaders enable developers of Tripal sites to collect and store various types of data that are required to build comprehensive genomic and genetic databases. Tripal is a toolkit that combines a Chado database schema and Drupal, a content management system, for construction of online biological databases. Several data loaders available in Tripal, such as the GFF3, FASTA and OBO loaders, enable genome data to be loaded relatively easily into Chado. While these loaders fulfill the needs for many common file formats, MCL meets the long-standing need for data loaders for a variety of other important data types. This presentation provides details on the data templates, user interface and the admin interface of MCL. This module is implemented in the Genome Database for Rosaceae (rosaceae.org), CottonGEN (cottongen.org), Citrus Genome Database (citrusgenome.org), Genome Database for Vaccinium (vaccinium.org) and the Cool Season Food Legume Database (coolseasonfoodlegume.org). MCL and user documentation are available for download at https://github.com/tripal/mainlab_chado_loader/releases/latest.

P0046: Methods: Bioinformatics

Tracking Top 20 Associations from Four Whole Genome Association Study (GWAS) Programs with Varied Input Data Quantity

Yan Yan, University of Saskatchewan, Saskatoon, SK, Canada

Genome-wide association studies (GWAS) have become an effective tool in plant breeding and genetics. GWAS programs link genotypes to phenotype using statistical models. One important influence on the statistical reliability of these models is the amount of input data supplied. The objective of this project is to investigate how the strongest associations reported by four popular GWAS programs change with varying amounts of data input.
Data from a previous *Arabidopsis* study containing 391 genome accessions (up to 3 replicates each) and 1100 phenotypes were used. Data subsets from 10% to 90% of the original were created using Monte-Carlo sampling. To de-emphasize possible sampling effects and biases, two sampling strategies were used. Genotype data were filtered for minor allele frequency < 0.05, matched to phenotype subsets, and run through GWAS programs PLINK, TASSEL, GAPIT and FaST-LMM. The original top 20 associations from each program were compared to the strongest 20 output by the same program using reduced subsets for both sampling strategies.
All programs showed approximately linear trends between input data and the number of original top 20 associations retained. However, PLINK showed a pronounced deviation between 60% and 70% of data used. Results were heavily influenced by sampling strategy. For example, half of the original top associations were no longer retained at 80% of original data with one strategy, and at 50% with the other strategy. Given the costs involved in producing input data for GWAS, this project provides valuable input regarding the number of genomes required for meaningful results.

P0047: Methods: Bioinformatics
ORCAE: Online Resource for Community Annotation of Eukaryotes
Lieven Sterck[1], Thomas Van Parys[2], Stephane Rombauts[3] and Yves Van de Peer[3], (1)Ghent University - VIB, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium, (2)VIB-UGent, Ghent, Belgium, (3)VIB Ghent University, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium
Conducting gene and genome annotation typically relies on diverse information resources going from sequence to expression data depending on whether structural or functional annotation is performed. To help researchers doing gene annotation while having access to these different data types, we developed ORCAE (Online Resource for Community Annotation of Eukaryotes), a web-technology-compliant portal for use in community genome annotation efforts.
ORCAE allows browsing and on the fly editing of gene structures and descriptions, moreover all manual curations are immediately visible for other users and for each locus a history of modifications is available.
Through its interface, ORCAE offers easy access to precomputed information that greatly facilitates the work of a curator. The gene page offers several informative graphics with a focus on gene structure quality (eg. multiple alignments of homologs) helping human annotators in improving the proposed automated annotation. Annotators can then use the build-in GenomeView interface to easily check/modify gene structures. A unique feature of ORCAE is that the portal is highly dynamic: on modification of a gene model, all the available information is immediately updated and presented on the gene page.
ORCAE can both be used to coordinate ongoing annotation efforts as well as to present published genomes to the public by acting as a genome portal. Therefore it is equipped with all the necessary features: advanced text-search, Blast functionality and a genome browsing interface (AnnoJ).
ORCAE is available at http://bioinformatics.psb.ugent.be/orcae/ and can now also be downloaded for local installation.

P0048: Methods: Bioinformatics
GEMT: Unified GWAS for Genetic and Environmental Interaction and Multiple Trait Model
Jiabo Wang, Washington State University, Pullman, WA and Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA
Association study is becoming the major method to map genes controlling complex diseases and agriculturally important traits. These traits are usually correlated and measured under multiple environments on individuals with certain structure and relationship. These multiple dimensional covariance makes computation extremely time consuming. The practical analyses were forced to simplify or partially ignore the existence of the correlations, which cause inflated false positives and reduction of statistical power. Upon the new development of BLINK (http://zzlab.net/blink), we developed an algorithm that consider all the relationship and developed a computationally affordable software package, GEMT, to conduct Genome Wide Association Study (GWAS) with either genetic and environmental interaction, or multiple trait model. Key features of GEMT include: 1) simultaneously considering causes for both correlation among individuals and and correlations among environments or traits; 2) computationally tractable, e.g. a dataset with 2500 individuals, 614K markers, and four environments, only take three hours on a Mac Pro with 12 cores. The GEMT package, including executable program, user manual, and demo data, are available at http://zzlab.net/GEMT.
Keywords: Gene mapping, Association study, genetic and environmental interaction, multiple variates.

P0049: Methods: Bioinformatics
High-Resolution Haplotype Construction by Low-Coverage Single Gamete Cell Sequencing
Ruidong Li[1], Mikeal L. Roose[1], Sergio Pietro Ferrante[1], Han Qu[2] and Zhenyu Jia[2], (1)University of California, Riverside, CA, (2)University of California Riverside, Riverside, CA
A haplotype is a set of DNA variants (or alleles) that inherits together from a single parent. Increasing evidences have shown that phenotypic variations can be better explained if haplotype data are available. Unrevealing the haplotype structure is also critical for understanding allele-specific events such as asymmetric methylation on DNA strands and parent-specific de novo mutations. Many computational methods have been proposed for haplotype inference. However, these methods either require genotypes of both parents of the individual which sometimes are not available, or need a large sample from the individual's progeny (>100) which are costly and time-consuming. Genotyping single haploid cells (e.g., gametes) from a heterozygous individual provides opportunities for an efficient reconstruction of the parental haplotypes because the complexity of mixtures of paternally and maternally inherited DNA in these cells has been substantially reduced. We have developed a novel and efficient algorithm for high-resolution haplotype construction using DNA sequence data of a small number of haploid cells at low coverage. Both simulation study and real gamete sequence data analysis showed that the new algorithm outperforms other competing methods.

Our method dramatically reduces the required sample size and sequencing depth for haplotype phasing, thus, will greatly facilitate whole-genome haplotype association studies and other large-scale haplotype or chromosomal recombination relevant research.

P0050: Methods: Bioinformatics

Comparison of Various Gene Prediction Pipelines to Generate Evidence-Based Gene Models for *Cocos nucifera* var. Catigan Green Dwarf (CATD) Genome Annotation

Charles Anthon E. Cadorna[1], Patrick Michael V. Chiuco[1], John Rafael A. Ferrer[1], Hideki Nogochi[2], Shinji Kondo[2] and Cynthia P. Saloma[1,3], (1)Philippine Genome Center, University of the Philippines, Quezon City, Philippines, (2)National Institute of Genetics, Mishima, Japan, (3)National Institute of Molecular Biology and Biotechnology, University of the Philippines, Quezon City, Philippines

The agronomic importance of coconut, as well as its lengthy growth period before reproduction has led to the sequencing of its genome to enable genomics-assisted breeding. A high-quality draft genome assembly of *Cocos nucifera* var Catigan Green Dwarf (CATD) has already been produced, however, a set of high-quality gene models is not yet available. Here, we present a comparative analysis of various gene prediction pipelines to generate high-quality gene models for the structural and functional annotations of *Cocos nucifera* var. Catigan Green Dwarf (CATD) draft genome. *Cocos nucifera* var. Catigan Green Dwarf (CATD) draft genome assembly version 3.0, with assembled sequences representing ~100% of the 2.9 Gb-long genome, was used to predict gene models. We have compared the gene prediction results of Seqping, BRAKER1, and MAKER-P for *Cocos nucifera* var. Catigan Green Dwarf (CATD). Similarities and variation in the prediction result of each tool is evident during comparative analysis. The completeness of the annotated gene space was evaluated using BUSCO (Benchmarking Universal Single-Copy Orthologs) and CoreGF.

P0051: Methods: Bioinformatics

Standardizing Gene Names in Key Vertebrate Species

Bethan Yates, Susan Tweedie, Bryony Braschi, Paul Denny, Kristian Gray, Ruth Seal and Elspeth Bruford, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Standardized gene nomenclature provides an essential resource for all researchers. The Vertebrate Nomenclature Committee (VGNC), operating in parallel with the HUGO Gene Nomenclature Committee (HGNC), was established in 2016 to approve consistent gene names and symbols across vertebrate species that lack their own nomenclature group. Our naming strategy for each species starts by identifying a high confidence set of genes with consistently predicted 1:1 human orthologs. These orthologs are assigned the human gene nomenclature using an automated pipeline. This strategy has resulted in >10K genes being named in each of chimpanzee, cow, dog and horse. We are now focussing on naming the non-consensus orthologs, many of which belong to complex gene families and require careful manual review across multiple species. One such example is the keratin gene family where only around half of keratin orthologs were named via our automated pipeline; our manual curation of this family extends the recently published catalog of equine and canine keratin genes (Balmer et al., PLoS One 2017) to include bovine genes. VGNC nomenclature data are available via our online vertebrate gene nomenclature portal (vertebrate.genenames.org) and are displayed on other key resources including Ensembl and NCBI Gene. We will continue to add new species to VGNC based on the quality of genome assembly and annotations, perceived importance as a model for humans and demand from the research community. Please email us if you have expertise in a particular species or gene family you could help us to name vgnc@genenames.org

P0052: Methods: Bioinformatics

Genomic Utilities of BUSCO v3

Felipe A. Simão, Robert M. Waterhouse, Mathieu Seppey, Panagiotis Ioannidis, Evgenia V. Kriventseva and Evgeny M. Zdobnov, University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland

The advent of high-throughput genomics has brought about a veritable paradigm shift in biological research. Due to rising demands and increasing volumes of data, technologies and downstream analysis tools have been rapidly evolving. This makes thorough quality control of the 'products' of sequencing data, e.g. genomes, genes, or transcriptomes, essential. Addressing this need, the Benchmarking Universal Single-Copy Orthologues (BUSCO) assessment tool provides intuitive quantitative measures of genomic data completeness in terms of expected gene content (Simão et al, 2015, PMID:26059717, http://busco.ezlab.org). BUSCO assessments identify complete, duplicated, fragmented, and missing genes and enable like-for-like quality comparisons of different datasets. These features mean that BUSCO has rapidly become established as an essential genomics tool, using up-to-date data from many species and with broader utilities than the popular but now discontinued Core Eukaryotic Genes Mapping Approach (Parra et al, 2007, PMID:17332020). Selected from major species clades (amongst prokaryota and eukaryota) of the OrthoDB catalog of orthologs, 44 clade-specific datasets can be used with BUSCO v3, permitting analysis using a large number of highly specific single-copy genes across all domains of life. Here we present a summary of the latest BUSCO features along with a variety of scenarios highlighting the wide range of uses of BUSCO assessments, designed primarily for (i) performing genomics data quality control, but also applicable for (ii) building robust training sets for gene predictors, (iii) selecting high-quality reference strains or species for comparative analyses, (iv) identifying reliable markers for large-scale phylogenomics studies, and (v) separating haplotypes in highly heterozygous assemblies.

P0053: Methods: Bioinformatics

Rare Splice Sites in Plant Protein-Coding Genes

Shengqiang Shu, David M. Goodstein and Daniel S. Rokhsar, DOE Joint Genome Institute, Walnut Creek, CA

We implemented an intron scoring scheme based on matches to consensus sequences of splice sites and their respective branch point and incorporated it into PASA to allow rare splice site introns supported by transcriptome in protein-coding gene annotation. Using rare splice site protein-coding genes as homology seeds, we predicted rare splice site genes from 62 Viridiplantae genomes using exonerate. Predicted rare splice sites are verified by aligning respective transcriptome short reads to CDS sequences. Some genes with rare splice site genes are

conserved widely among Angiosperm genomes while a few are specific to a clade. Based on sequences of splice sites and best scoring branch point in 30 conserved genes, ATAC intron in these genes is likely handled by U2-type spliceosome.

P0054: Methods: Bioinformatics

G-OnRamp: Create Genome Browsers for Collaborative Eukaryotic Genome Annotations

Yating Liu, Washington University in St. Louis, St. Louis, MO

G-OnRamp (http://gonramp.org) is a collaboration between two successful and long-running projects — the Genomics Education Partnership (GEP; http://gep.wustl.edu) and the Galaxy Project (https://galaxyproject.org). G-OnRamp provides biologists with an integrated, web-based, scalable environment for interactive annotation of eukaryotic genomes using large genomic datasets. It also provides educators with a platform to help undergraduates develop "big data" science skills through eukaryotic genome annotation.

GEP is a consortium of over 100 colleges and universities that provides Classroom Undergraduate Research Experiences (CURE) in bioinformatics/genomics for students at all levels. GEP faculty currently use the annotation of multiple *Drosophila* species to introduce genomics and research thinking to undergraduates. Galaxy is a popular open-source, web-based scientific gateway for accessible, reproducible, and transparent analyses of large biomedical datasets. G-OnRamp extends Galaxy with tools and workflows that creates UCSC Assembly Hubs and Apollo/JBrowse genome browsers with evidence tracks for sequence similarity, *ab initio* gene predictions, RNA-Seq, and repeats. Educators can use this system to design CUREs based on their favorite eukaryotic species (*e.g.*, parasitoid wasps).

G-OnRamp provides a VirtualBox virtual appliance and an AMI image for local and cloud (Amazon EC2) deployments. Future versions of G-OnRamp will (i) enable data storage with CyVerse; (ii) support additional configuration options for UCSC Assembly Hubs; and (iii) adapt GEP annotation tools for other informant species. We will host G-OnRamp training workshops in June and July 2018 at Washington University in St. Louis. If you are interested in attending, please sign up for the G-OnRamp mailing list at http://gonramp.org/signup. Supported by NIH 1R25GM119157.

P0055: Methods: Bioinformatics

A Statistical Framework for Detecting Mislabeled and Contaminated Samples using Shallow-Depth Sequence Data

Ariel W. Chan, Cornell University, Ithaca, NY

Researchers typically sequence a given individual multiple times, either re-sequencing the same DNA sample (technical replication) or sequencing different DNA samples collected on the same individual (biological replication) or both. Before merging the data from these replicate sequence runs, it is important to verify that no errors, such as DNA contamination or mix-ups, occurred during the data collection pipeline. Methods to detect such errors exist but are often ad hoc and require some combination of genotype calling, imputation, and haplotype phasing, making them unsuitable for error detection in low- to moderate-depth sequence data where such tasks are difficult to perform accurately. Additionally, because most existing methods employ a pairwise-comparison approach for error detection rather than joint analysis of the putative replicates, results may be difficult to interpret. We introduce a new method for error detection suitable for shallow-depth sequence data. Using Bayes Theorem, we calculate the posterior probability distribution over the set of relations describing the putative replicates and infer which of the samples originated from an identical genotypic source. Our method addresses key limitations of existing methods and produced highly accurate results in simulation experiments. We examined the impact of read depth, the number of sites analyzed, and allele frequency on algorithmic performance, using both real and simulated data. Our method is implemented as an R package called BIGRED, which is freely available for download: https://github.com/ac2278/BIGRED.

P0056: Methods: Bioinformatics

Gymno-PLAZA: An Access Point for Comparative Genomics in Conifer Species

Lieven Sterck, Ghent University - VIB, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium, Zhen Li, VIB-UGent, ghent, Belgium, Yves Van de Peer, VIB - Ghent University, Zwijnaarde, Belgium; VIB Ghent University, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium and Klaas Vandepoele, Center for Plant Systems Biology, VIB, Ghent, Belgium; Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

Comparative sequence analysis has significantly altered our view on the complexity of genome organization and gene functions in different kingdoms. Gymno-PLAZA is designed to make comparative genomics data for conifers available through a user-friendly web interface. Structural and functional annotation, gene families, protein domains, phylogenetic trees, and detailed information about genome organization can easily be queried and visualized. The functional annotation has been updated and now comprises data from Gene Ontology, MapMan, UniProtKB/Swiss-Prot, PlnTFDB and PlantTFDB. Furthermore, we included improved algorithms to transfer functional annotation from well-characterized plant genomes to other species. These data and features make gymno-PLAZA (http://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/) a versatile and comprehensible resource for users wanting to explore genome information to study different aspects of conifer biology, both in model and non-model organisms.

P0057: Methods: Bioinformatics

An Issue for Practical Use in Genomic Prediction of Trait Segregation in a Progeny Population

Shiori Yabe, Institute of Crop Science, NARO, Ibaraki, Japan; JST, PRESTO, Saitama, Japan and Takeshi Hayashi, Institute of Crop Science, NARO, Tsukuba, Japan

Selection of good parental combination which can generate progeny with high genetic performance is important in cross breeding for plant species. Genomic prediction can be applied to prediction of trait segregation in a progeny population. to decide the combination of parents to be crossed, where selection of parental combinations would be made such that mean genotypic values is high and/or some individuals with extremely high genetic performance are emerged in the progeny. Therefore, the prediction models, which are constructed in a training population of parents to be crossed, are required to accurately predict the genetic variance as well as the mean genetic value in the progeny. In the present study, we evaluated the effects the prediction methods on the prediction of genetic variance in a progeny population with simulation experiments. Here, we used two prediction models obtained by genomic best linear unbiased prediction (GBLUP) and least absolute shrinkage

and selection operator (LASSO). Both prediction methods could accurately predict the mean genotypic value of a progeny population. However, genetic variance of a progeny population was underestimated in GBLUP, suggesting the shrinkage of estimated effects may result in the underestimation of genetic variance. LASSO could predict it better than GBLUP, although LASSO sometimes overestimated the variance especially when the number of QTL was large. Thus, the balance between the number of QTL and markers with effects much affect the accuracy in predicting genetic variance in a progeny population.

P0058: Methods: Bioinformatics

Knowledge Representation and Database Integration to Facilitate Genetic Analysis and Development of Underutilised Crop Plants

Liliana Andres, Southern Cross Plant Science, Southern Cross University, Lismore, Australia

A massive amount of genetic, trait and associated experimental data have been generated in recent decades for major crops. In order to facilitate access and data sharing, the scientific community has emphasized the necessity to adopt controlled vocabularies. However, this requires information to be transformed into systematic and structured terms, which are designed to classify large quantities of information. Computational biologists have proposed solutions to this problem by developing "ontologies", structured vocabularies within specific domains, enabling an explicit representation of their properties and the relations between them. It is critical nowadays to establish ontology databases for a wide range of crops, in order to improve the outcomes from accumulated research not only for major crops, but also emerging underutilized crops. For example, Bambara groundnut (*Vigna subterranea*) a drought and rainfall tolerant crop, and an important source of protein and nutrition for a large sub-Saharan African population. This project aims to organize information related to underutilised crops, with the grain legume bambara groundnut (*Vigna subterranea*) as an exemplar, focusing on seed nutrient composition. The approaches developed will enable data sharing that may contribute to the improvement of Bambara and other underutilised crops to benefit from research investment in related major crop taxa.

P0059: Methods: Bioinformatics

Preparation of the Database for the *Ds* Transposon Tagging Lines of *Arabidopsis thaliana* from RIKEN BRC

Satoshi Iuchi and Masatosmo Kobayashi, RIKEN Bioresource Center, Tsukuba, Japan

RIKEN BRC (BioResource Center) has been joined the National Bioresource Project (NBRP) at Japan and the Experimental Plant Division of RIKEN BRC has been distributed the mutants lines of Arabidopsis thaliana to the researcher around the world. The Ds lines induced mutation by inserting maize Ds transposon element using the Ac/Ds transposable element tagging system into the Arabidopsis genome, accession Nossen. Here we present the database for Arabidopsis mutant lines, Ds transposon tagging lines.

To accelerate the research which uses the Ds lines, we try to contract the database and web site. At first, we mapped the DNA sequence from Ds border sequences to Arabidopsis genome sequence using CLC workbench. The Columbia genome sequence was used for this mapping as mapping reference, because the genome sequence of Nossen accession genome has not determined yet. After mapping, we construct the database include both of these mapped position data and Arabidopsis gene annotations from Araport11 and TAIR10. To construction for web site, we try to make visualize the data at simple web page. In this purpose, this web page has been displayed a schematic view of the insert region. Now we try to improve the web page design for more easy to get information about Ds lines (Web site address; https://plant.rtc.riken.jp/resource8/ds_line/ds_line_list.html)

P0060: Methods: Bioinformatics

Practical Molecular Classification Unit (IPSUM) to Define Species Border of Eukaryote and to Friendly Augment Linnaeus Classification

Ikuo Nakamura, Graduate School of Horticulture, Chiba University, Matsudo, Japan

Classification of species has been a base of various areas in biological researches. Eukaryotic species have been classified according to their morphological characteristics and described using Linnaeus' binomial nomenclature. Because recent tremendous advance of sequencing technology, complete genomic sequence of type or representative strain has been resolved in more than 400 eukaryotic species. Therefore, type strain of a species can be identified using molecular analysis. However, there is no molecular-based classification system that can define the boundary of a species that is very important to reveal molecular mechanism of species speciation.

We found that a particular protein tag (Ptag) sequence (*ca*. 370 aa) within the largest subunit (POLA1) of RNA polymerase I complex showed species-specific variation among 4000 eukaryotic species. Any Ptag sequence was searchable with homology from local-blast database of 4000 Ptag sequences. This indicated that the Ptag sequence was unique and rapidly differentiated during the speciation. Ptag sequences of more than 100 lines belonging to the same species, such as *Homo sapiens*, *Arabidopsis thaliana*, *Oryza sativa*, and *Saccharomyces cerevisiae*, were found to be highly conservative within a species. Here we propose identical Ptag sequence unit member (IPSUM), which is a basic molecular classification unit sharing identical Ptag sequence. Depending on species, one IPSUM corresponds to one Linnaeus species, or plural IPSUMs may correspond to one species, or vice versa.

The classification system using Ptag sequence is very simple but will be a useful to identify every eukaryotic species and to friendly augment Linnaeus' classification.

P0061: Methods: Bioinformatics

Haplotype Inference using SNP Data in Polyploid Potato

Johan H. Willemsen, Richard G.F. Visser and Herman J. van Eck, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands

We developed a new haplotype inference algorithm for polyploids, to reconstruct haplotypes on the basis of using un-phased genotype calls from either GBS data or SNP arrays. Existing software for polyploid haplotyping, such as ShesisPlus, Satlotyper and polyhap, lack the ability to process the large amounts of SNPs as present in highly heterozygous polyploid crops such as potato.

The major improvements of our approach rely on a novel approach for joining short haplotype segments (estimated with the EM algorithm), allowing to efficiently characterize haplotypes using genotype calls from SNP array data.
Our results show that our approach is able to process un-phased SNP data, and reconstruct haplotypes with reasonable accuracy. We use our approach to successfully obtain phased haplotypes of a set of amplicon sequences. In addition we explored the usage of our haplotype inference method on genotypic data from the SOL-STW SNP array.

P0062: Methods: Bioinformatics
Exploiting Short Read Sequencing for the Characterization of Haplotypes in Autotetraploid Potato
Johan H. Willemsen, Richard G.F. Visser and Herman J. van Eck, Wageningen University & Research, Plant Breeding, Wageningen, Netherlands
The gene pool of potato varieties is characterized by a large number of alleles. Previous studies have identified 10-25 distinct alleles for many genes, and each tetraploid variety can have up to four different alleles per locus. So far routine application of haplotype diversity in tetraploids is challenging due to the lack of reliable software to reconstruct haplotypes.
We investigated the use of short read next-generation sequencing to adequately reconstruct haplotypes in the tetraploid potato, and developed a novel method for reconstructing (partial) haplotypes.
We divide the haplotype reconstruction into two steps: Short-range haplotype reconstruction, followed by haplotype extension. With a simulation study we demonstrated that this approach results in highly accurate haplotypes, but we also verified the approach with resequencing data of tetraploid potato cultivars. Moreover we highlight the challenges and limitations of using haplotype assemblies y for the reconstruction of haplotypes in tetraploid potato.

P0063: Methods: Bioinformatics
***De novo* Optical Map Assembly: A Signal-Based Approach**
Dick de Ridder, Wageningen University, Wageningen, Netherlands
Optical maps (OM) are restriction fragment length maps of DNA molecules. Single molecule optical maps can technically be as long as the DNA strand extracted from a sample. Leading in high-throughput OM is Bionano Genomics (BNG), with its Irys and Saphyr platforms, based on a nick-labeling protocol and DNA nanochannel technology. Nick-labeling involves using an enzyme to nick a single DNA strand by removing several nucleotides at a specific hexa-nucleotide sequence. This strand is subsequently ligated using fluorescently labelled nucleotides and run through nanochannels, while obtaining fluorescence images of the DNA strands. The OM data produced on BNG platforms has two main inconsistencies: 1. due to a high labeling error rate, false negative and false positive labels occur frequently (around 15%); 2. due to the flexibility of DNA molecules, the distance between the same nicks may vary between different OM molecules. This renders the OM assembly phase a challenge. Currently, the only OM assembler for high-throughput single molecule OM data is delivered by BNG, but this suffers from a number of shortcomings. First, around half of the OM molecule data is discarded due to quality filtering; second, it is closed source, which prevents inspecting or improving the underlying algorithms. Therefore, we have developed a tool called PhoTOMap (Pixels To Optical MAP), to improve assembly efficiency by extracting and processing raw molecule signals. Obtained fluorescence signals are then used in a novel de novo OM assembly approach, PhoTOAsm, which directly makes use of these signals rather than their translations to sequences.

P0064: Methods: Bioinformatics
Development of a Gene Family Toolkit for Exploring Diversity in New Sequence Data
Paul C. Bailey[1], Ritesh Krishna[2], Burkhard Steuernagel[3], Brande B.H. Wulff[3], Alistair Darby[4] and Anthony Hall[1], (1)Earlham Institute, Norwich, United Kingdom, (2)International Business Machines, Manchester, United Kingdom, (3)John Innes Centre, Norwich, United Kingdom, (4)University of Liverpool, Liverpool, United Kingdom
As more genomes are sequenced it is becoming apparent that gene duplication and deletion are important drivers in evolution, with rapidly expanding gene families often a signature of their role in an organism's adaptation to the environment. To identify these events, we are building a gene family analysis toolkit which will be deployed on the Cyverse cloud infrastructure for use by the scientific community. Central to its purpose will be the ability to distinguish evolutionary relationships between genes within gene families for currently and newly sequenced species, both at the intra- and inter-species level.
As a pilot study, we are using a collection of landrace bread wheats (the Watkins collection) which have been sequenced by exome capture to explore the diversity of the large Nucleotide Binding Leucine Rich Repeat (NLR) family of plant resistance genes in the collection. Illumina read data from each wheat line have been assembled using various tools. The resulting contigs have been aligned to their corresponding subgroups within the NLR family. Looking at subgroups that are expanded relative to other monocot species, we can demonstrate that further novel gene duplication events have occurred in specific lines of the Watkins collection. The next step will be to understand whether specific genes in the family are under positive selection and therefore which genes have particular functional significance. The assembly and downstream procedures will be placed into a Docker container for use on Cyverse as a tool for exploring the diversity of any gene family in any species with sequence data.

P0065: Methods: Bioinformatics
GAIA: Integrated Metagenomics Suite
Andreu Paytuvi Gallart, Ermanno Battista, Fabio Scippacercola, Walter Sanseverino and Riccardo Aiese Cigliano, Sequentia Biotech, Bellaterra, Spain
Identifying the biological diversity of a microbial population is of fundamental importance due to its implications in industrial processes, environmental studies and clinical applications. Today, there is still a need to develop new, easy-to-use bioinformatics tools to analyze both shotgun and targeted metagenomics with the highest accuracy and the lowest running time. With the aim of overcoming this need, we introduce GAIA, an online Software as a Service (SaaS) solution that has been designed to provide the maximum information from whatever metagenomics sample: 16/18S, virome or shotgun analysis. GAIA is able to obtain a comprehensive and detailed overview at any taxonomic

level of microbiomes of different origins: human (e.g. stomach or skin), agricultural and environmental (e.g. land, water or organic waste). Recent publications have benchmarked commonly-used 16/18S pipelines (Siegwald, *et al*. 2017) as well as shotgun metagenomics pipelines (McIntyre, *et al*. 2017), and we also benchmarked GAIA with the same datasets. GAIA is currently the best pipeline to analyze shotgun metagenomics data as it obtained the highest F-measures above all tested pipelines (CLARK, Kraken, LMAT, BlastMegan, DiamondMegan, NBC and OneCodex). In addition, GAIA also obtains excellent F-measures analyzing 16S data, yielding better F-measures than CLARK, kraken, BMP, mothur and QIIME. The overall objective of GAIA is to provide to academia and industries with an integrated metagenomics suite that will allow to perform metagenomics data analysis easily, quickly, and affordably with the best accuracy.

P0066: Methods: Bioinformatics

New Insights into the Analysis of Hi-C Data

Andreu Paytuvi Gallart[1,2], Riccardo Aiese Cigliano[1], Walter Sanseverino[1], Covadonga Vara[2,3] and Aurora Ruiz-Herrera[2,3], (1)Sequentia Biotech, Bellaterra, Spain, (2)Departament de Biologia Cel·lular, Fisiologia i Immunologia, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, (3)Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Hi-C is a genome-wide approach that allows to study the chromatin structure by identifying any contact between a pair of loci by coupling proximity-based ligation with deep sequencing (Erez Lieberman-Aiden, *et al*. 2009). Current bioinformatics approaches to analyze Hi-C data are resource-hungry besides of failing at the time to deliver the results within a reasonable timeframe and to perform visual or statistical analysis of two or more contact maps. Here we describe HiCloud, a bioinformatics pipeline integrated to a web framework that only needs the upload of the reads and few clicks to obtain: heatmaps, TADs, compartments, and differential interacting bins if more than two conditions are provided. In terms of speed, the pipeline has been benchmarked together with other pipelines, such as HiCUP (Wingett, *et al*. 2015) or TADbit (Serra, *et al*. 2017), using 10 CPUs and a subset of 5M Hi-C reads from real data generated in our lab coming from mouse cells. The time consumed by HiCloud to complete mapping, filtering and preparation of the results was ~14m. In contrast, HiCUP and TADbit required ~52m and ~10h, respectively. In terms of performance, HiCloud have 155.415 more valid reads than TADbit, 113.722 out of them (73.1%) due to higher mapping efficiency. The overall objective of HiCloud is to provide to academia and industries with an online integrated tool that will allow to perform Hi-C data analysis easily, quickly and affordably, without the need to have bioinformatics skills or powerful machines.

P0067: Methods: Bioinformatics

AIR: Artificial Intelligence RNASeq

Andreu Paytuvi Gallart, Ermanno Battista, Fabio Scippacercola, Walter Sanseverino and Riccardo Aiese Cigliano, Sequentia Biotech, Bellaterra, Spain

In the field of genomics, sequencing technologies have drastically changed in the last few years and the output of complex data generated has outpaced the solutions available for analysis, integration and interpretation. RNA Sequencing has emerged as the number one technique in transcriptomics and thus the solution we propose is based on this. A.I.R.: Artificial Intelligence RNASeq is the first easy to use SaaS (Software as a Service) built with solid scientific methods. AIR is able to perform a robust DEG and GEOA analysis with different statistics to solve three important obstacles in the genomics field simultaneously: the informatics problem (specifically data storage, automatization of results and duration of analysis); the scientific problem (data interpretation and data integration, as well as providing new bioinformatics and statistical functions); the social problem (the lack of availability of skilled bioinformaticians). The overall objective of this project is to introduce a disruptive innovation that will allow researchers to perform transcriptomics data analysis easily, quickly and affordably. AIR is accessible at http://transcriptomics.cloud

P0068: Methods: Bioinformatics

Tissue-Specific Pathway Comparison in Mammalian Organisms

Nadezhda T. Doncheva[1], Oana Palasca[2], Lars J. Jensen[3] and **Jan Gorodkin**[2], (1)Center for non-coding RNA in Technology and Health & Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark, (2)Center for non-coding RNA in Technology and Health, Department of Veterinary and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark, (3)Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark

A very important part of studying human diseases and developing of new treatment therapies is the use of animal models. In order to generate reliable hypotheses about a phenotype of interest in human, we need to overcome the challenge of identifying the regulatory genes and pathways in an animal model that would be of interest. Moreover, it is crucial to take into account the intrinsic differences in regulation between human and animal models as well as the mechanisms underlying the specific phenotype.

Thus, we have designed a unique comparison framework for investigating pathways in mammalian organisms on different levels of detail based on data integration and network analysis techniques. The orthology relationships between genes in different organisms as indicated by the eggNOG database (http://eggnogdb.embl.de) were used to identify and visualize the conservation of already existing pathways. We showed that all KEGG pathways overlap more than 50% between human, mouse, rat and pig and even half of them overlap at least to 75%. Furthermore, we integrated the pathways with tissue expression data from the recently updated TISSUES database (http://tissues.jensenlab.org/), which now covers several mammalian organisms. The obtained orthology pathways open for the possibility of identifying tissue expression differences and similarities, which can point to parts of the pathways related to a human disease of interest and thereby suggest the most suitable organism for modeling this disease.

P0069: Methods: Bioinformatics

Rapid Pathogen Identification Pipeline Based on the Analysis of Single Nucleotide Polymorphism

Chul Lee[1], Sojin Lee[1] and Heebal Kim[2], (1)IPBI, Seoul National University, Seoul, South Korea, (2)Seoul National University, Seoul, Korea, Republic of (South)

Conventional typing methods (e.g., pulsed-field gel electrophoresis) are constantly substituted by new microbial subtyping methods based on the next generation sequencing (NGS) data with rapid advances in whole genome sequencing (WGS) technologies. Especially, the single nucleotide polymorphism (SNP)-based comparative genomic and phylogenetic methods which compare sequences to reference genome has been proposed, and some pipelines have already been applied to recent food poisoning outbreak investigations. It is important to continue putting efforts in developing better pipelines which perform more quickly and accurately. Since food poisoning requires a rapid action, possible confusion resulting from wrong SNP detection must be avoided. In this regard, we constructed much easier and accurate pathogen subtyping pipeline, **SNP** identification for Strain Typ**ing(SNPing)** for foodborne pathogens. In order to validate performance of our newly developed pipeline program, simulated *Salmonella enterica* genome data sets which contains known SNPs, insertions and deletions information were created and tested for comparing accuracy and analytical time. In testing, SNPing pipeline was measured at 99% sensitivity and 99.9% accuracy for SNP detection and showed 30% reduction in analytical time than the existing pipeline program. Tests using fifty-five salmonella genome data with outbreak information obtained from NCBI database were also performed to demonstrate if SNPing is applicable to the actual cases as well, and with comparison to the outbreak information, it was appeared to classify the sample data accurately. Based on the results, it indicates that SNPing is an appropriate WGS-based subtyping method which can be applied in the food poisoning pathogen identification.

P0070: Methods: Bioinformatics

Microservices for Phytozome

Joseph W. Carlson, Patrick Davidson, Richard D Hayes and David M. Goodstein, DOE Joint Genome Institute, Walnut Creek, CA

The Phytozome project (https://phytozome.jgi.doe.gov/) is the JGI's compendium of landplant assemblies, annotations and analyses consisting of genomes sequenced by the JGI, collaborators and community members. Our web infrastructure uses a combination of a customized databases architecture and community standard databases (CHADO, InterMine and BioMart) for providing the datastore to deliver data. Using large, monolithic databases such as InterMine or BioMart has posed some problems in our project because of the time and effort needed to load and preprocess the data. Since the Phytozome project is constantly upgrading its database to include new organisms, the extra time for data loading has delayed some updates.

We have been experimenting with replacing some of our monolithic database and tomcat-based servlet architecture with lighter weight microservice components. Rather than delivering data from one source, we fragment across customized databases with simple node.js based web servers.

These services include:

1) A server which provides a interface to our collection of VCF files for a dataset of natural popul
ation variations. This service leverages the ability of PostgresQL to store and search json records
which captures all information in a VCF file.

2) A server which connects to a referential CHADO database for sequence and structure of our assemblies, annotations and analyses.

3) A server for performing pairwise protein blast sequences on demand. Rather than precompute and store pairwise alignments between homologous proteins as we have in the past, alignments will be generated on demand.

P0071: Methods: Bioinformatics

Understanding the Liver Under Heat Stress with Statistical Learning: A Multiomics Computational Approach

Allen Hubbard[1], Xiaoke Zhang[2], Abhi Singh[1], Sara F. Jastrebski[1] and Carl J. Schmidt[3], (1)University of Delaware, Newark, DE, (2)George Washington University, Washington, DC, (3)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Obtaining biological insight from large-scale multi-omics data is challenging due to biological and technical variance. Careful experimental design can limit unwanted noise. However, when properly harnessed, heterogeneity can be used to detect biological signals that elude traditional enrichment analysis. For example, biological variation relating to a treatment response depends on many variables that are not easily controlled such as allelic or physiological variants. This fact can be informative because many compounds involved in the same process will have similar patterns of heterogeneity. This can be used to identify relationships between elements of the same pathway, even when their scales of expression and variance differ considerably, by relying on statistical learning strategies. This approach allows the combination of transcriptome and metabolome data to gain a more comprehensive biological understanding of a system. We assemble a pipeline using these strategies to identify genes and metabolites that are strong biomarkers for the heat stress response in the broiler chicken liver. Then, we develop models for the identified pathways that revolve around altered dynamics of carbon through mechanisms of antioxidant and lipid production.

P0072: Methods: Bioinformatics

Data-Driven Discovery for Drivers of the Effects of Diet on Disease

Richard Linchangco, Jeremy J. Jay and Cory Brouwer, University of North Carolina at Charlotte, Kannapolis, NC

Epidemiological studies have observed that plant-based diets confer human health benefits. The molecular mechanisms underlying the health benefits of plant-based foods remain largely unknown. The scientific literature contains latent, informative relationships connecting entities of diet and disease that aid in explaining these mechanisms. High throughput methods in data-mining are capable of extracting the relationships from literature, but result in an explosion of candidate relationships between diet and disease entities.

An approach for filtering and determining the relevance of the millions of mined relationships from scientific literature is needed to generate hypotheses about the molecular effects of diet on disease.

An NLP-based text-mining software was used in this study to extract relationships between diet and disease entities from abstracts in MEDLINE and Agricola. Mined relationships were integrated with curated public databases to generate a diet-disease network. We developed a weighted association rule mining approach which extends the Kulczynski null-invariant measure to determine meaningful relationships in our diet-disease network that describe the molecular effects of dietary components on the human genome. As a case study for our approach, we selected *Brassica oleracea var. italica*, broccoli, and analyzed our results to identify phytochemicals and genes related to human disease development.

This study presents an association rule mining approach for weighting and ranking text-mined relationships that aid in explaining the molecular mechanisms by which diet affects human health. This work also provides a diet-disease network for data-driven hypotheses generation in the support of nutrigenomics and personalized nutrition questions.

P0073: Methods: Bioinformatics

Integrated Transcriptomics and Metabolic Pathway Analysis in Plants using Open-Science Cyberinfrastructure of KBase

Vivek Kumar[1], Sunita Kumari[1], Jim Thomason[1], Doreen Ware[2], Priya Ranjan[3], Sean McCorkle[4], Samuel M. D. Seaver[5], Nomi Harris[6], Robert W. Cottingham[3], Christopher Henry[5] and Adam Arkin[6], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)USDA/ARS - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)Oak Ridge National Laboratory, Oak Ridge, TN, (4)Brookhaven National Laboratory, Upton, NY, (5)Argonne National Laboratory, Argonne, IL, (6)Lawrence Berkeley National Laboratory, Berkeley, CA

The U.S Department of Energy Systems Biology Knowledgebase (KBase, http://kbase.us) is an open, web-accessible system for systems biology research focused on microbes, plants and their communities. It provides a range of integrated biological data types and associated analysis tools (Apps) that include modeling, simulation methods and visualizations. Scientifically, KBase Apps currently include genome assembly and annotation, gene expression analysis, metabolic modeling, comparative genomics and microbial community analysis. The KBase Software Development Kit enables developers in the KBase user community to add new apps to the system.

KBase has a rich set of computational methods and curated datasets for gene expression analysis based on RNA-seq, including a selection of preprocessed high-quality reference genomes and a wide variety of algorithms for short-read mapping, identification of splice junctions, differential expression analysis, and visualization. KBase supports both the original and new Tuxedo tool suites, including Bowtie2, TopHat2, HISAT2, Cufflinks, StringTie, Cuffdiff, Ballgown and DESeq2. KBase also provides services that are compatible with gene expression profiles for downstream analysis, including clustering of expression profiles based on different algorithms and comparison of reaction fluxes with gene expression values to identify metabolic pathways where expression and flux data agree or conflict.

KBase services are available from within an interactive, Jupyter-based user interface that supports the creation of dynamic workflow documents called Narratives that enable experimental and computational biologists to work together to share and publish data, approaches, workflows, and conclusions, leading to transparent and reproducible computational experiments.

P0074: Methods: Bioinformatics

Accurately Dissecting Genetic Effects for the Analysis of Complex Plant Traits

Wenchao Zhang[1], Bongsong Kim[1], Xinbin Dai[1], Shizhong Xu[2] and **Patrick X. Zhao**[1], (1)Noble Research Institute, Ardmore, OK, (2)Department of Botany & Plant Sciences, University of California, Riverside, CA

Complex plant traits, such as yield, tolerance to abiotic and biotic stresses, are often governed by many individual genes (G), the gene-gene interactions (GxG) and gene-environment interactions (GxE). Therefore, analysis of complex plant trait demands accurately dissecting these genetic causal effects, i.e. accurately identifying individual genes (loci), gene-gene and gene-environment interactions that are consistently associated with the observed phenotypes. In this report, we describe our recently developed a trio of genotype-phenotype association analysis tools, namely 1) GWASPRO (bioinfo.noble.org/GWASPRO/), which adopts a simple linear mixed model (LMM) for the analysis of *additive* genetic effects (referred as one-dimensional mapping or 1D GWAS here) and is specially optimized for the analysis of "big data" generated from large-scale genome-wide association studies (GWASs); 2) PEPIS (bioinfo.noble.org/PolyGenic_QTL/), which adopts a full polygenic linear mixed model to analyze the *additive* (1D GWAS)*, dominance effects* (1D GWAS) and *epistatic effects* such as *additive* x *additive*, *additive* x *dominance, dominance* x *additive, dominance* x *dominance* (referred as two-dimensional mapping or 2D GWAS here) in GWASs and quantitative trait loci mapping; and 3) PATOWAS (bioinfo.noble.org/PATOWAS/), which further extends the 2D GWAS LMMs for broader associative 'omics' studies, i.e. the LMMs can not only be applied to GWASs, but also transcriptomics-wide association studies (TWASs) and metabolomics-wide association studies (MWASs). Our case analysis of a set of publically available Immortalized F2 (IMF2) associative 'omics' studies, which consisted of comprehensive GWASs, TWASs and MWASs, demonstrated the high performance of our developed LMMs and tools, enabling genotype-phenotype association discovery and genetic variances analysis for complex plant traits.

P0075: Methods: Bioinformatics

ArrayExpress and Expression Atlas: Tools for Archiving, Searching and Visualizing Functional Genomics Data at EMBL-EBI

Suhaib Mohammed, Laura Huerta, Nuno Fonseca, Maria Keays, Amy Tang, Elisabet Barrera, Wojtek Bazant, Anja Fullgrabe, Nancy George, Haider Iqbal, Alfonso Munoz-Pomer, Irene Papatheodorou, Alvis Brazma and Robert Petryszak, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Expression Atlas at EMBL-EBI (https://www.ebi.ac.uk/gxa) is a database and web-service that provides information about gene and protein expression patterns in plant and animal species across different tissues, developmental stages and diseases. All datasets are curated with a semi-automatic process of identifying the experimental factors, such as diseases or perturbations, annotating metadata with Experimental Factor Ontology terms (EFO) and describing the experimental comparisons for further processing. Analyses of RNA-seq datasets are performed using our standardized pipeline iRAP (https://nunofonseca.github.io/irap) while our microarray pipeline use standard open source tools. Expression Atlas provides baseline and differential studies. Baseline studies report transcript abundance within tissues, developmental stages or cell lines while differential studies report changes in expression across two different conditions, for example, healthy versus disease. Presently, we provide results on more than 3,100 experiments from more than 30 species. Large studies include BluePrint, GTEx, Encode, CCLE, HipSci and Pan-Cancer. A quarter of all studies are plant experiments and 11% are relevant to diseases. All data in Expression Atlas are free to browse, download, and reuse.

ArrayExpress at EMBL-EBI (https://www.ebi.ac.uk/arrayexpress) is the source of most data in Expression Atlas. It is an archive for functional genomics experiments, supporting their reuse by the research community. Experiments in ArrayExpress are directly submitted by scientists

through Annotare (https://www.ebi.ac.uk/fg/annotare/login/), our webform-based tool. Accession numbers are generated within 15 minutes of submission, pre-published data sets can be kept private, and submitter's identity can be hidden for double-blind review.

P0076: Methods: Bioinformatics

A Rapid Epistatic Mixed-Model Association Analysis by Linear Retransformations of Genomic Estimated Value

Chao Ning[1], Dan Wang[1], Huiming Kang[1], Raphael Mrode[2], Lei Zhou[1], Shizhong Xu[3] and **Jian-Feng Liu**[1], (1)China Agricultural University, Beijing, China, (2)International Livestock Research Institute, Nairobi, Kenya, (3)Department of Botany & Plant Sciences, University of California, Riverside, CA

Motivation: Epistasis provides a feasible way for probing potential genetic machanism of complex traits. However, time-consuming computation challenges successful detection of interaction in practice, especially when linear mixed model (LMM) is used to control type I error in the presence of population structure and cryptic relatedness.

Results: A rapid epistatic mixed-model association analysis (REMMA) method was developed to overcome computational limitation. This method first estimates individuals' epistatic effects by an extended genomic best linear unbiased prediction (EG-BLUP) model with additive and epistatic kinship matrix, then pairwise interaction effects are obtained by linear retransformations of individials' epistatic effects. Simulation studies showed that REMMA could control type I error and increase statistical power in detecting epistatic QTNs in comparison with existing LMM-based FaST-LMM. We applied REMMA to two real datasets, a mouse dataset and the Wellcome Trust Case Control Consortium (WTCCC) data. Application to the mouse data further confirmed the performance of REMMA in controlling type I error. For the WTCCC data, we found most epistatic QTNs for type 1 diabetes (T1D) located in a major histocompatibility complex (MHC) region, from which a large interacting network with 12 hub genes (interacting with ten or more genes) was established.

Acknowledgments: We appreciate the financial support from the National Major Development Program of Transgenic Breeding (2014ZX0800953B).

P0077: Methods: Bioinformatics

Animal QTLdb and CorrDB Updates: Integrative Development of Genetics/Genomics Databases and Tools to Meet New Challenges

Zhiliang Hu, Iowa State University, Iowa, IA, Carissa Park, Iowa State University, Ames, IA and James M. Reecy, Department of Animal Science, Iowa State University, Ames, IA

Since the inception of the Animal Quantitative Trait Loci Database (QTLdb) and the Animal Trait Correlation Database (CorrDB), development of database infrastructure and tools has steadily evolved to meet user needs and the demands of ongoing growth in the amount and types of data. To date, 136,137 QTL/association data on 1,890 different traits in 6 species have been curated from 1,881 publications in 14 years (as of August, 2017). A total of 3,635 correlation data between 276 different traits in cattle have been used to establish a database model linking the two databases, in an effort to expand livestock genetic/genomic information networks. New developments include CorrDB curator tools integrated with those of QTLdb, and a better structured environment for sharing of trait ontology development resources and centralized curator activity management. The co-development of the two database platforms allows information transfer from genetically/phenotypically correlated traits to their QTL/association locations in the genome, and further to curated genome features such as genes, SNPs, and other types of variations. A new trait modifier creation tool has extended the capacity to manage trait variants in a scalable structure. One of the major goals in the development of these database resources is to facilitate more organized, inclusive, and complete data curation, toward a well-structured "big data" reservoir and improved future utilization of the data. We welcome your suggestions regarding database improvements to ensure these resources maintain their value and usefulness to the community.

P0078: Methods: Bioinformatics

Evaluating the Haplotype Phasing of *de novo* Assemblies using FALCON Unzip

Brett Hannigan, DNAnexus, Mountain View, CA and Chai Fungtammasan, DNAnexus

Advances in long-read sequencing technology have allowed researchers to produce de-novo genome assemblies of remarkable quality and continuity. Today, it is not uncommon for new genomes to be assembled with N50 contig lengths that easily exceed a megabase in length, prompting many researchers to brag about their inclusion in the "1Mb Contig Club" or even the "10 Mb Contig Club." As these longer contig assemblies become more common, attention is now being given to the ability to form a "phased" assembly – that is the ability to correctly group sequence that comes from the same parental allele on the same contig. To this end, PacBio recently released their FALCON Unzip program to create regions of phase haplotypes from a FALCON assembly. In this talk, we evaluate the performance of FALCON Unzip in forming these phased haplotypes by assembling and phasing the genomes of a macaque and human. By examining SNP's that are known to be unique to one of the parents, we show that FALCON Unzip is able to produce impressive phasing information requiring nothing more than a little additional time in the compute environment to process the data.

P0079: Methods: Bioinformatics

Using a Combination of Sequencing Methods for Improved *de novo* Genome Assembly

Alana J Firl[1], Nestor Kippes[1], Eric Dowd[2], Luca Comai[1] and Isabelle M. Henry[1], (1)University of California, Davis, CA, (2)Wm. Wrigley Jr. Co.

With the advent of many new sequencing technologies, reference sequences are rapidly being developed for a variety of model and non-model species. Our goal was to evaluate whether and how combining different sequencing methods could improve the *de novo* assembly of a 500Mb, heterozygous, diploid genome. Specifically, we used sequence reads from PacBio RSII, Chicago and HiC libraries (Dovetail Genomics), 10X Genomics, and optical mapping (Bionano) and compared the assembly quality of each method to the qualities of hybrid assemblies that used different combinations of these methods. We found that, of the individual assembly methods, the PacBio FALCON assembly at 60X coverage produced the draft genome with the highest N50 and closest in size to the expected 500Mb genome. We used our optical mapping data to assess the quality of the different assemblies obtained either with individual technologies or combinations of technologies. Finally, we

combined the PacBio reads, Dovetail assembly, and optical mapping to resolve misjoins and close gaps for further improvement of the assembly.

P0080: Methods: Bioinformatics

(Soybean-PIPE): A Computational Approach in Soybean Functional Genomics

Bahram Samanfar[1], Elroy Cober[1], Stephen Molnar[1], Brad Barnes[2], James Green[2], Frank Dehne[3] and Ashkan Golshani[4], (1)Agriculture and Agri Food Canada, Ottawa Research and Development Centre, Ottawa, Ottawa, ON, Canada, (2)Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, Canada, (3)Department of Computer Sciences, Carleton University, Ottawa, ON, Canada, (4)Department of Biology, Carleton University, Ottawa, ON, Canada

Soybean is one of the major Canadian grain crops and its production is expanding in Canada with the majority of the increase in short season areas (Western Canada and northern regions). So far, more than eleven maturity loci have been reported in soybean, however, the molecular basis of almost half of them is not yet clear. The list of novel factors affecting these pathways in soybean, and in model plants like Arabidopsis, continues to grow suggesting the presence of other novel players which are yet to be discovered.

The soybean Protein-protein Interaction Prediction Engine (Soybean-PIPE) is a computational tool used to predict protein-protein interactions (PPI) in soybean. Protein-Protein Interactions (PPIs) are essential molecular interactions that define the biology of a cell, its development and its responses to various stimuli. Theoretically, if a gene interacts with groups of genes involved in one specific pathway, that gene might also be involved in that specific pathway ("guilt by association"). Our knowledge of global PPI networks in complex organisms such as human and plants is restricted by technical limitations of current methods.

Briefly, PIPE searches for re-occurring short polypeptide sequences between known interacting protein pairs and novel proteins and predicts interactions based on protein sequence information and a database of known interacting protein pairs. PIPE has been used to produce proteome-wide, all-to-all predicted interactomes in a variety of organisms including yeast (*Saccharomyces cerevisiae),* human (*Homo sapiens),* Arabidopsis and others. PIPE is typically tuned, for a given organism, to achieve a specificity of 99.95%, i.e., to be conservative in predicting PPIs. PIPE has been independently evaluated and compared to other PPI prediction methods and has been shown to significantly outperform the others in terms of recall-precision across all of the datasets tested. It has also been shown that PIPE has the ability to produce cross-species predictions, Soybean-SCN and Soybean-Human.

Currently we are using PIPE towards predicting the first comprehensive protein-protein interaction network for soybean ever generated. In an independent study (Samanfar et al., 2017), we have used three different approaches; bioinformatics (Soybean-PIPE), classical plant breeding, and molecular biology (analysis of SSR and SNP haplotypes) to identify a novel gene involved in time of flowering and maturity in soybean. This strategy successfully identified a new maturity locus tentatively called "E10" and the underlying candidate gene (FT4).

Identification of molecular markers tagging the PIPE-identified genes controlling flowering and maturity in soybean will allow soybean breeders to efficiently develop varieties using molecular marker assisted breeding. Allele specific markers will allow stacking of early maturity alleles to develop even earlier maturing cultivars. This bioinformatics approach (soybean-PIPE) will also help to bridge the gap in knowledge of the flowering and maturity pathway in soybean and can be applied to other important traits such as seed protein content, oil quality and host-pathogen interactions (Soybean-SCN).

P0081: Methods: Bioinformatics

Variant Calling Quality Enhancement within a Large Breeding Population

Jan van Oeveren, Keygene, Wageningen, Netherlands, **Bas Tolhuis**, GENALICE, Harderwijk, Netherlands, Antoine Janssen, KeyGene, Wageningen, Netherlands and Roeland van Ham, KeyGene N.V., Wageningen, Netherlands

Whole genome sequencing applications are rapidly being adopted in agrigenomics to improve molecular breeding strategies. This massively increases data production and creates growing computational and storage challenges in a time and budget constraint environment.

As a solution to this bottleneck in NGS analysis KeyGene has implemented GENALICE MAP as secondary analysis software for ultra-fast mapping and variant calling of complex crop genomes. Processing times have been reduced over a 100 fold, whereas former hardware requirements of a cluster >1000 cores decreased to a single machine with 12 cores. The acceleration thus achieved enables large cohort, multi sample, variant calling for population genomics applications.

In this study the use, accuracy and efficiency of the GENALICE Population Caller is characterized for SNP variation of a large breeding population of highly homozygous individuals derived from complex crosses between multiple parental lines. This population set-up allows for the assessment of true genotype composition from parental haplotype block information. The data set is assessed with individual line and population based variant scores. We demonstrate the quality enhancement capabilities of population calling procedures over a single sample calling process, such as more accurate variant calls, identification of new variants and reductions in false positive and false negative calls.

P0082: Methods: Bioinformatics

Machine Learning Identifies Unique Features of Reproductive Phased, Secondary, Small Interfering RNAs in Rice and Maize

Parth U. Patel[1], Sandra Mathioni[2], Atul Kakrana[2], Hagit Shatkay[1] and Blake Meyers[2], (1)University of Delaware, Newark, DE, (2)Donald Danforth Plant Science Center, St. Louis, MO

Work in the field of plant small RNAs (sRNA) (21 to 24 nt) has identified numerous components of the RNA interference machinery, playing roles in biological processes such as development, epigenetic modifications, and plant defense. In recent years, a class of endogenous plant sRNA called *phased, secondary, small interfering RNAs* (phasiRNAs) have been identified in plants. These are generated from a long RNA precursor in phase of 21- to 24-nt. We and others have shown that maize anthers (male reproductive organs) express two classes of phasiRNAs (21- and 24-nt) during different developmental time points. Other data demonstrate that perturbation of these phasiRNA can disrupt male fertility. Our study focuses on grass-specific reproductive phasiRNAs in maize and rice. To characterize these small RNAs, we (1) optimized a supervised machine learning based classification pipeline to sort previously unknown sequences from plants to find reproductive phasiRNAs from among other types of small RNAs, (2) observed position-specific biases in phasiRNAs relative to other sRNAs, and (3) illustrated the

duplex nature of phasiRNA and other sRNA biogenesis impacting nucleotide composition as well as AGO-sorting. Future work will include the analysis of putative targets and proposed functions of reproductive phasiRNAs. Ultimately, the knowledge gained from this unique set of features from the phasiRNAs and the other sRNAs may help us understand both their biological roles and their evolution.

P0083: Methods: Bioinformatics

Nutritional Epigenomics: Evaluating the Impact of Maternal Methionine Supplementation on Fetal Developmental Programming using Multi-Omics Data

Rocio Amorin, University of Florida, Gainesville, FL

Maternal nutrition can induce modifications in the epigenome of the fetus, such as DNA methylation, which can alter gene expression and impact future performance. This is related to nutrition via the one carbon cycle, involving methionine. We hypothesized that an increased dietary methionine in beef cows during early gestation will impact the epigenome and transcriptome of the offspring. Muscle samples from 20 bull calves (10 maternal control diet and 10 maternal methionine-rich diet) were collected at 1 month of age. Both whole-genome DNA methylation and gene expression were evaluated through next-generation sequencing methods. Out of 1.5M cytosines evaluated, 36k were differentially methylated (DMC's) between treatments (qvalue < 0.05). 4.5k genes, defined as differentially methylated (DMG's), had 5 or more DMC's within or near (±10kb) their genic region. A gene set analysis revealed that these DMG's are involved in pathways such as calcium signaling pathway, ion channel activity and transporting activity. We analyzed the genome in 1kb sliding windows, and found that 0.5% of the windows had at least 3 DMC's, suggesting that DMC's are not uniformly distributed. At least 30 of these windows were identified as CpG islands, and were in close proximity to histone genes, transcription factors, and genes involved in gene expression regulation. The RNA-Seq analysis identified 13k genes expressed in muscle. Our findings provide further evidence that maternal diets in early gestation can shape the epigenome of fetal tissues, which in turn may lead to permanent phenotypic changes in the phenotype with lifelong consequences.

P0084: Methods: Bioinformatics

polyRAD: Genotype Calling with Uncertainty from Sequencing Data in Polyploids and Diploids

Lindsay V. Clark, Alexander E. Lipka and Erik J. Sacks, Dept of Crop Sciences, University of Illinois, Urbana, IL

All genotypes derived from genotyping-by-sequencing or RAD-seq protocols have inherent uncertainty associated with them. Heterozygotes can be miscalled as homozygotes due to undersampling of alleles, homozygotes can be miscalled as heterozygotes due to sample contamination, and in polyploids, allele dosage in heterozygotes may be ambiguous without very high read depth. Passing this uncertainty on to downstream analysis can increase the power to detect associations between genotype and phenotype, in comparison to simply using the most likely or most probable genotype, which may be incorrect. polyRAD is a new R package that uses Bayesian statistics to estimate genotype probabilities based on allelic read depth and population parameters. Read depth can be imported from reference and non-reference pipelines, including TASSEL, GATK, Stacks, and TagDigger. Population structure is accounted for when estimating genotype prior probabilities, increasing accuracy over similar algorithms that assume Hardy-Weinberg Equilibrium. Genotype probabilities can also be estimated for many types of mapping populations. polyRAD can model inheritance in both autopolyploids and allopolyploids, and allows the inheritance mode itself to be uncertain. Future versions of the algorithm will incorporate linkage disequilibrium, eliminating the need for imputation after genotype calling. Genotypes are exported as a mean of all possible genotypes, weighted by genotype probabilities, in order to be used directly by the software GAPIT or rrBLUP for genome-wide association and genomic prediction.

P0085: Methods: Bioinformatics

A New Tool to Prioritize Candidate Genes and Characterize Sample Behavior in Differential Expression Analysis of Transcriptomic Data

Danielle Novick, Michael Babak, M. Joseph Tomlinson IV, Jose Daniel Chazi Capelo and Behnam Abasht, University of Delaware, Newark, DE

High-throughput sequencing technologies and differential expression analysis tools have facilitated the profiling of differential gene expression, resulting in sometimes extensive lists of candidate genes. However, prioritization of these candidate genes remains a challenge, especially in non-model systems where data mining techniques cannot easily be used to inform the analysis. As a result, many researchers use fold-change (FC) and p-value to prioritize genes, which does not account for the ability of individual samples to skew the results. ARB (Analysis of RNA-seq Data Behavior) addresses this problem by classifying genes based on the degree of separation of expression values between experimental groups. This is accomplished with a supervised subsetting algorithm that incrementally excludes samples up to a cutoff set by the user to achieve non-overlapping separation between groups. If a subset can be assembled that meets the user's constraints, then the following is calculated: (i) inter-group separation index (IGSI); 0-2 scale where a value of 2 corresponds to a subset where no samples have been removed (ii) FC using only that subset of samples, and (iii) the identities of the samples that have been excluded from the subset. With this information, a researcher can easily select the genes with the highest IGSI and FC for further scrutiny. ARB also includes functions to identify outliers, rank samples based on relative expression across genes, and identify misclassified samples based on frequency of exclusion. The program is designed to run using output from Cuffdiff and is available with sample data at https://github.com/AbashtLaboratory/ARB.

P0086: Methods: Bioinformatics

Mining DNA Methylation Variations in Alleles and Homeologs using CGmapTools

Weilong Guo, China Agricultural University, Beijing, China

DNA methylation is important for gene silencing and imprinting in both plants and animals. We developed software CGmapTools (freely available at https://cgmaptools.github.io/) as a toolset for mining DNA methylation information in BS-seq data, by integrating ~40 command-lines applications into one package. This package uses CGmap and ATCGmap as the format interfaces, and designed binary formats to reduce the file sizes and support fast data retrieval, and can be applied for context-wise, gene-wise, bin-wise, region-wise, and sample-wise analyses and visualizations. To accurately identifying heterozygous SNVs from partially C-to-T converted, we designed two methods, BayesWC and BinomWC, that substantially improved the precision of heterozygous SNV calls from ~80% to 99% while retaining comparable recalls. With

these SNV calls, we provided functions for allele-specific DNA methylation (ASM) analysis and visualizing the methylation status on reads. Applying ASM analysis to a previous dataset, we found that an average of 1.5% of investigated regions showed allelic methylation, which were significantly enriched in transposon elements and likely to be shared by the same cell-type. A dynamic fragment strategy was utilized for DMR analysis in low-coverage data. Recently, we develop new method for mining differential DNA methylations among homeologs, suitable for allopolyploid genomes, such as bread wheat. The new method also support visualising DNA methylomes variations and genomic variations among homeologs.

P0087: Methods: Bioinformatics

ELIXIR Belgium: Tools and Services

Frederik Coppens, VIB, Gent, Belgium, Lieven Sterck, Ghent University - VIB, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium, Klaas Vandepoele, Center for Plant Systems Biology, VIB, Ghent, Belgium and Yves Van de Peer, VIB - Ghent University, Zwijnaarde, Belgium

The Belgian ELIXIR node will be instrumental to optimize the interactions between the national level and ELIXIR-Europe, and for promoting consistency and complementarity between the services offered by ELIXIR Belgium and the other European nodes. ELIXIR Belgium will focus on upgrading the current node services and expanding the portfolio of node services, in the fields of human health and sustainable agriculture. Furthermore, trainings and workshops related to data science will be offered.

P0088: Methods: Bioinformatics

TripalMap Mapviewer 1.0

Katheryn Buble, Sook Jung, Jodi L. Humann, Chun-Huai Cheng, Taein Lee, Stephen P. Ficklin, Jing Yu and Dorrie Main, Washington State University, Pullman, WA

TripalMap MapViewer provides an interactive and responsive visualization for genetic maps using the D3.js Javascript technology. Similar in functionality to the existing GMOD-CMap map comparison tool, TripalMap offers the benefit of using map data directly stored in GMOD-Chado, a generic database schema. Data duplication can be reduced for sites that maintain both Chado and CMap as no separate underlying database is required with TripalMap. Also TripalMap is an extension of Tripal, integrating directly with the Tripal API framework for online site visualization of genomic, genetic and breeding data. The TripalMap interface can be integrated in any Tripal map page and hyperlinked from any Tripal page that is displayed in maps (marker, QTL, heritable morphological marker and/or gene). MapViewer 1.0 displays all map linkage groups and zooming for specific regions. Genetic marker comparison between two linkage groups along with Correspondence Matrix and Dot Plot tools is supported and future versions will offer comparison between multiple linkage groups. A control panel provides configuration for marker and QTL colors and display patterns. Materialized Views allow customization providing both better performance and flexibility in the way data is stored in Chado. Future versions will provide further configuration for marker and QTL colors and display patterns in the Admin page, and enhanced filtering for marker correspondences and comparison of entire maps. TripalMap MapViewer is available for use at Cool Season Food Legume Database (https://www.coolseasonfoodlegume.org/MapViewer) and source code: https://github.com/tripal/tripal_map

P0089: Methods: Bioinformatics

Genome Variation Map: A Repository of Genome Variations for Global Animals and Plants

Shuhui Song, Beijing Institute of Genomics, Beijing, China and Zhang Zhang, BIG Data Center, Beijing Institute of Genomics, Beijing, China

The Genome Variation Map (GVM; http://bigd.big.ac.cn/gvm/) is a public data repository of genome variations. As a core resource in the BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, GVM dedicates to collect, integrate and visualize genome variations for a wide range of animals and plants, accepts submissions of different types of genome variations from all over the world and provides free open access to all publicly available data in support of worldwide research activities. Unlike existing related databases, GVM features integration of a large number of genome variations for a broad diversity of species including human, cultivated plants and domesticated animals. Specifically, the current implementation of GVM not only houses a total of ~4.9 billion variants for 19 species including chicken, dog, goat, human, poplar, rice and tomato, but also incorporates 8,669 individual genotypes and 13,262 manually curated high-quality genotype-to-phenotype associations for non-human species. In addition, GVM provides friendly intuitive web interfaces for data submission, browse, search and visualization. Collectively, GVM serves as an important resource for archiving genomic variation data, helpful for better understanding population genetic diversity and deciphering complex mechanisms associated with different phenotypes.

P0090: Methods: Bioinformatics

GenSAS v5.1: A Web-Based Platform for Structural and Functional Genome Annotation and Curation

Jodi L. Humann[1], Taein Lee[2], Stephen P. Ficklin[1], Chun-Huai Cheng[2], Heidi Hough[1], Sook Jung[1], Jill L. Wegrzyn[3], David B. Neale[4] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (4)Department of Plant Sciences, University of California, Davis, Davis, CA

The Genome Sequence Annotation Server v5.1 (GenSAS, www.gensas.org) is a web-based annotation and curation platform that combines several common annotation tools into one easy-to-use, integrated resource. The user-friendly interfaces, with embedded instructions, guide users through the annotation process. GenSAS has annotation tools for eukaryotes and prokaryotes and supports model and non-model organisms. Users can upload a variety of evidence files to support the annotation process for their genome sequence. These include GFF3 files of aligned features and previous annotations; FASTA files of repeat, transcript, EST, or protein sequences; and gene models from Genbank. GenSAS also allows users to upload Illumina RNA-Seq reads, align the reads to the genome using TopHat, and use the data to train the gene model prediction program Augustus, which allows for more accurate gene models for eukaryotic genomes, especially non-model organisms. JBrowse and Apollo are integrated into GenSAS allowing structural annotation results to be easily viewed and manual curation to be

performed. Users can share GenSAS projects with other users enabling collaborative or community wide curation. GenSAS also has a functional annotation step to assign protein functions and identify functional domains for the official gene set. After the annotation process is complete, the final step of the GenSAS pipeline generates the required files for publication which includes merging the manual annotations from Apollo into the final annotation.

P0091: Methods: Bioinformatics
iPat: Intelligent Prediction and Association Tool for Genomic Research
Chun-Peng Chen, Washington State University, Pullman, WA
The ultimate goal of genomic research is to effectively predict phenotypes from genotypes so that medical management can improve human health and molecular breeding can increase agricultural production. Genomic prediction or selection (GS) plays a complementary role to genome-wide association studies (GWAS), which is the primary method to identify genes underlying phenotypes. Unfortunately, most computing tools cannot perform data analyses for both GWAS and GS. Furthermore, the majority of these tools are executed through a command-line interface (CLI), which requires programming skills. Non-programmers struggle to use them efficiently because of the steep learning curves and zero tolerance for data formats and mistakes when inputting keywords and parameters. To address these problems, this study developed a software package, named the Intelligent Prediction and Association Tool (iPat), with a user-friendly graphical user interface (GUI). With iPat, GWAS or GS can be performed using a pointing device to simply drag and/or click on graphical elements to specify input data files, choose input parameters, and select analytical models. Models available to users include those implemented in third party CLI packages such as GAPIT, PLINK, FarmCPU, BLINK, rrBLUP, and BGLR. Users can choose any data format and conduct analyses with any of these packages. File conversions are automatically conducted for specified input data and selected packages. A GWAS-assisted genomic prediction method was implemented to perform genomic prediction using any GWAS method such as FarmCPU. iPat was written in Java for adaptation to multiple operating systems including Windows, Mac, and Linux.

P0092: Methods: Bioinformatics
Mapping 1000000 Markers with Flipper
Charles F. Crane, USDA-ARS, West Lafayette, IN
Flipper is a fast linkage and deletion mapper that implements Kruskal's algorithm to produce a minimum spanning tree of genetic markers. Additional heuristics allow Flipper to recognize and repair many instances of misplaced markers and misjoined subsets of markers. Flipper requires memory in proportion to the square of the number of markers. With 16 Gb of system memory, it cannot run more than about 43000 markers at once without paging to disk. Flipper allows efficient deduplication of marker sets to mutually recombining exemplar markers, and this can be used to reduce 1000000 perfectly genotyped markers to a sufficiently small number of exemplars. Since Flipper keeps track of the nonrecombining markers that accompany each exemplar, those can be inserted into the map based on exemplars. However, this method fails with even 0.1% genotyping error, since the errors introduce spurious recombination and greatly increase the number of apparent exemplars. A better, more general method is to map a subset of enough markers to place ca. 100 per linkage group, and then use this map as a set of mathematical attractors to cluster the entire marker set into linkage groups that might fit within the 43000-marker limit. Since each recombination fraction is calculated only once, this grouping of markers requires very little memory. The same idea can be applied to abutting subsets of markers within a single linkage group. The markers that fall into each non-overlapping cluster can then be mapped and the clusters can be sutured together with local permutation of 8-10 markers at each junction. An example of each of these three methods is given for a simulated genotyping-by-sequencing experiment with 1000000 markers.

P0093: Methods: Bioinformatics
Metaomgraph for 'Omics Data: NoSQL-Enabled Big Data Visualization and Analysis
Eve Syrkin Wurtele, Manhoi Hur and Urminder Singh, Iowa State University, Ames, IA
MetaOmGraph (MOG) is a tool for plotting and analyzing large sets of data while using as little memory as possible. It was designed with transcriptomic data in mind, but is data-type agnostic. Curated compilations of RNA-seq datasets, including from Arabidopsis, maize, yeast and humans, most composed of thousands of samples, are available on our website. Alternately, a researcher can analyze her/his own dataset. Features include: visualizing gene expression patterns; sorting data by any metadata terms; finding groups of genes with common functions; determining which genes have expression patterns most/least correlated to a gene of interest; statistical determinations of significance. We will highlight MOG function using RNA-Seq and metabolomics datasets from maize and yeast.

P0094: Methods: Bioinformatics
James Hutton Institute Informatics Visualization Software
Iain Milne, Gordon Stephen, Paul Shaw, Sebastian Raubach and David Marshall, The James Hutton Institute, Dundee, United Kingdom
Modern day genomics, genetics and plant breeding relies on high-throughput sequencing and phenotyping technologies, generating data sets that have been increasing over the last few years in terms of both size and complexity. Storage and analysis of these diverse data sets is only possible via a combination of utilizing data warehousing, HPC clusters (often via the parallelization of existing code sets), information visualization, and visual analytics technologies.
At the James Hutton Institute we develop novel software tools, web applications, databases and information resources (https://ics.hutton.ac.uk/software) which allow users to explore and query their data in logical and intuitive ways – ultimately leading to improved solutions for scientific data management and information dissemination.
Germinate 3 stores various diverse data types and acts as a hub for other tools: Flapjack is built around graphical genotyping enabling users to sort and manipulate lines based on their genotype or on observed or predicted phenotypes; CurlyWhirly is a 3D viewer for PCA/PCo data; Tablet provides 2nd-generation sequence assembly and alignment visualization; and Helium utilizes plant pedigrees as a visualization framework. These resources are used by many institutes, companies, and large international projects, including the Genomic & Open-source

Breeding Informatics Initiative (GOBII), Seeds of Discovery (including UK Seed), Crop Wild Relatives, and the International Wheat Yield Partnership.
We will describe our tools, and the benefits they offer, and show how they're evolving to embrace new technologies such as the Plant Breeding API (BrAPI) for connectivity with external tools and resources, and Galaxy for the pipelining of analyses/HPC utilization.

P0095: Methods: Bioinformatics

PiRATE: A Pipeline to Retrieve and Annotate Transposable Elements of Non-Model Organisms

Jérémy Berthelier[1], Nathalie Casse[2], Nicolas Daccord[3], Véronique Jamilloux[4], Bruno Saint-Jean[1] and Grégory Carrier[1], (1)IFREMER/PBA, Nantes, France, (2)Le Mans University/MMS, Le Mans, France, (3)INRA/IRHS, Beaucouzé, France, (4)INRA/URGI, Versailles, France

With the emergence of long reads sequencing, a renewed interest in repetitive sequences is ongoing. Among them, transposable elements (TEs) are mobile DNA sequences which constitute powerful drivers of genome evolution. An efficient strategy to conduct a *de novo* TE annotation is the following: TEs are detected from genomic data and automatically classified to construct a TE library, used to conduct the TE annotation. However, tools performing the classification step use databanks of known TEs and are less efficient to classify TEs belonging to poorly studied taxa. As example, we working on the Haptophyte *Tisochrysis lutea* and only 17 TE families are described for this phylum on RepBase (the most used TE database). In comparison 29,404 TE families are listed for the Metazoan. Facing this, we built a new bioinformatics pipeline named PiRATE. We optimized its detection step by using every existing TE detection approaches. The goal is to promote the full-length detection of every TE families, to facilitate their classification. PiRATE was controlled with genomic data of the model plant *A. thaliana* and is able to detect 80% of its TE families. With PiRATE we estimate that the genome of the Haptophyte *T. lutea* is constituted of 20.8% of TEs and that 3.8% represent putative mobile TEs. PiRATE is automated into a stand-alone Galaxy and is available through a virtual machine: http://doi.org/10.17882/51795

P0096: Methods: Bioinformatics

Lep-MAP3: Robust Linkage Mapping Even for Low-Coverage Whole Genome Sequencing Data

Pasi M Rastas, University of Helsinki, Institute of Biotechnology, Helsinki, Finland

Accurate and dense linkage maps are required for family-based linkage and association studies, quantitative trait locus (QTL) mapping, analysis of genome synteny and other analyses. Moreover, linkage mapping is one of the best tools to detect errors in *de novo* genome assemblies, as well as to anchor assembled contigs within chromosomes. Even a mapping cross of ten individuals will detect many assembly errors. With more individuals and markers, even more local errors can be detected and more contigs can be anchored. Linkage maps with more markers than recombinations have multiple markers at most map positions. This will anchor contigs more reliable by locally pinpointing each recombination. However, the tools that are currently available for linkage mapping are not well suited for very large number of markers nor individuals.
Here we present linkage mapping software Lep-MAP3, capable of analysing large datasets. It is fast and has small memory footprint, it can simultaneously analyse multiple families and requires little manual work and data curation. It can analyse low-coverage whole genome sequencing datasets on millions of markers and thousands of individuals. Such cost-efficient data enables comprehensive validation and refinement of genome assemblies. We demonstrate that Lep-MAP3 obtains very good performance already on 5x sequencing coverage and outperforms the fastest available software on accuracy and often on speed. We also construct *de novo* linkage maps with millions of markers on real 5-12x whole-genome sequencing data. Lep-MAP3 is freely available with the source code under GNU general public license from http://sourceforge.net/projects/lep-map3.

P0097: Methods: Bioinformatics

Comparative Analysis of Mammal and Angiosperm Phylogenomic Synteny Networks

Tao Zhao, Wageningen University, Wageningen, Netherlands

Comparative phylogenomic synteny (genomic context) analysis holds great promise for the inference of gene and genome evolutionary history. Utilizing the extensive available whole-genome resources, we have built complete microsynteny (local conserved gene order) networks for all genes of 87 mammalian and 107 angiosperms genomes, respectively. Thus, we can directly compare genome dynamics of these two major clades that have evolved and radiated during the last ~170 million years. To interpret the entire synteny network, we exploited network statistical parameters (i.e. average clustering coefficient, retention percentage, cluster sizes) to characterize and quantify various evolutionary features (i.e. conservation vs diversity) of gene families in a phylogenomic context. In addition, we dissected the composition and size distribution of all synteny clusters, which provide intriguing insights into the differing genomic architectures and dynamics of mammals and flowering plants. Sufficient representative genomes for synteny network construction in this study provide us clearer phylogenetic profiling patterns of synteny clusters. We will highlight several representative examples of lineage-specific clusters (i.e. unique genomic changes) that signal potential links between genomic context variation and the evolution of lineage-specific phenotypic traits.

P0098: Methods: Bioinformatics

Camoco: Identifying High Priority Candidate Genes from GWAS using Co-Expression Networks

Robert J Schaefer[1], Jean-Michel Michno[2], Elaine M Norton[1], Joseph R Jeffers[3], Owen Hoekenga[4], Brian P. Dilkes[5], Ivan Baxter[6], Molly E McCue[1] and Chad Myers[3], (1)University of Minnesota, St. Paul, MN, (2)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (3)University of Minnesota, Minneapolis, MN, (4)Cayuga Genetics Consulting Group, NA, NY, (5)Department of Biochemistry, Purdue University, West Lafayette, IN, (6)USDA-ARS/Donald Danforth Plant Science Center, St. Louis, MO

Camoco is a fully featured computational framework for building, analyzing and integrating gene co-expression networks with loci identified in genome wide association studies (GWAS). Hundreds of links between genetic markers (SNPs) and agro-economically important traits have been identified by GWAS. Yet, the causal gene or allele often remains unknown due to many genes being in linkage disequilibrium (LD) with

each of potentially dozens of genetic markers. Co-expression networks identify genes that share similar response patterns of gene expression making them a powerful tool for inferring the biological function of under-characterized genes. In the right biological context, sets of causal genes related to a GWAS trait will exhibit strong co-expression while inconsequential genes in LD with the marker exhibit random patterns of co-expression.
Camoco features methods to build, analyze, and explore co-expression networks using either microarray or RNA-Seq data. Once built, Camoco establishes a biological context for networks by evaluating their ability to recapitulate previously described ontologies (e.g. GO, KEGG, or MapMan). Vetted networks are then used to determine subsets of genes in close proximity to GWAS loci that are strongly co-expressed. GWAS SNPs are mapped to genes using a SNP-to-gene mapping algorithm using user-defined or map-based haplotype windows. High priority candidate genes are identified by evaluating gene-specific co-expression among candidate genes. Demonstrations will be shown using GWAS datasets and co-expression networks generated in both plants and animals. Camoco is free and open source software and available at http://github.com/LinkageIO/Camoco.

P0099: Methods: Bioinformatics
Mercator - A Fast Online Tool to Annotate Plant Genomes
Marie E Bolger, IBG-2, Forschungszentrum Jülich, Jülich, Germany, Rainer Schwacke, forschungszentrum jülich, Juelich, Germany and Bjoern Usadel, Forschungszentrum Jülich GmbH, Jülich, Germany
Genome sequencing has become a relatively standard practice in recent years and has resulted in circa 200 plant genomes already been sequenced. This necessitates that the downstream tools are able quickly and efficiently process these data. Mercator, an online gene annotation tool has recently been upgraded in response to this need. The high quality annotations together with the cluster upgrades enables gene function prediction for whole plant genomes and transcriptomes (30-150k genes) within minutes. This has been further improved by the addition of visualizations which allow users to immediately compare their results with reference genomes. This tool can be found at www.plabipd.de/portal/web/guest/mercator-ii-alpha-version-

P0100: Methods: Bioinformatics
***GEAUniversal*: A Web-Based Universal Gene Expression Atlas System for Managing, Analyzing and Sharing Large-Scale RNA-Seq-Based Transcriptome Data**
Xinbin Dai, Clarissa Boschiero, Zhaohong Zhuang and Patrick X. Zhao, Noble Research Institute, Ardmore, OK
We successfully developed *GEAUniversal*, which is a Web-based Universal Gene Expression Atlas System for Managing, Analyzing and Sharing Large-scale RNA-seq-based Transcriptome Data. The system is capable of hosting data from multi-species. *GEAUniversal* was implemented using Python, Flask and MySQL. The transcriptomic data are organized in hierarchical fashion, according to species, experiments, treatments and biological samples. *GEAUniversal* provides three core functions: 1) Expression Profile Query - to query the expression levels of genes in user-defined samples. We implemented an on-the-fly data normalization algorithm to enable querying gene expressions across experiments as the RNA-seq data normalization procedure is dependent upon user-selected dataset. 2) Differential Expression (DE) Analysis – to find differentially expressed genes between two samples or treatments using DESeq2 package. The returned differentially expressed genes can be filtered, sorted, and also displayed as bar or line charts. *GEAUniversal* can further identify enriched gene ontology (GO) terms in the list of user-defined genes from previous DE analysis. This function provides valuable insights to further identifying and analyzing biological pathways of the genes of interest. 3) Gene Co-Expression Analysis - to discover genes with similar expression pattern in user selected samples. The *GEAUniversal* also integrates a suite of scripts to simplify the system installation, including metadata and transcriptomic data population. To date, the *GEAUniversal* has been successfully deployed to empower several GEA projects, for example, the Gene Expression Atlas for Cultivated Alfalfa (*Medicago sativa*) at the Diploid Level (CADL) (www.alfalfatoolbox.org/atlasCADL/) and MtSSP-Atlas: A Gene Expression Atlas for Studying the Small Signaling Peptides in *Medicago truncatula* (mtsspdb.noble.org/atlas/).

P0101: Methods: Bioinformatics
CausNet: A Causal Inference Algorithm for Gene Regulatory Network Reconstruction
Xiaohan Kang[1], Faqiang Wu[2], Bruce Hajek[1] and Yoshie Hanzawa[2], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)California State University, Northridge, Northridge, CA
High-throughput sequencing has made large-scale transcriptional data under multiple experimental conditions available, yet gene regulatory network (GRN) reconstruction based on time-series data with statistical confidence is still a challenging task, partially due to limited data versus countless possible regulatory interactions, biological and technical variances, and multiple time scales of gene regulations. Reconstruction with confidence levels on the predicted regulatory interactions is essential for subsequent network analysis and experiment design.
We present CausNet, a causal inference algorithm that finds gene regulatory network reconstruction with confidence levels on the regulatory interactions. CausNet takes expression data and design matrix as input, and outputs a GRN with reliability scores that indicate the confidence levels. The algorithm is based on sparse linear regression model and Granger causality, where the former mitigates the limited data issue, and the latter removes indirect interactions. The reliability scores based on the biological replication data points are obtained by perturbation analysis, which is a Gaussian approximation of bootstrapping in statistics. The optimality of CausNet in the regression model makes it especially suitable for the study of a small number of core genes in a GRN.
We demonstrate the usage of CausNet to reconstruct clock gene networks of soybean and show its performance on simulated biologically plausible expression data. CausNet is implemented in Python 3 and is freely available at https://github.com/Veggente/soybean-network.

P0102: Methods: Bioinformatics
iDEP: Integrated Differential Expression and Pathway Analysis for RNA-Seq Data
Steven Xijin Ge, South Dakota State University, Brookings, SD

Analysis and interpretation of RNA-Seq data remain a challenge. We aim to develop a user-friendly web application for exploratory data analysis (EDA), differential expression, and pathway analysis. The key idea of iDEP (integrated Differential Expression and Pathway analysis) is to make many powerful R/Bioconductor packages easily accessible by wrapping them under a graphical interface, alongside annotation databases. For EDA, it performs hierarchical clustering, k-means clustering, and principal component analysis. iDEP detects differentially expressed genes using the limma and DESeq2 packages. For a group of co-expressed genes, it identifies enriched gene ontology (GO) terms as well as transcription factor binding motifs in promoters. Pathway analysis can be performed using packages like GAGE, GSEA, PGSEA, or ReactomePA. iDEP can also detect chromosomal gain or loss using the PREDA package. iDEP uses annotation of 69 metazoa and 44 plant genomes in Ensembl for ID mapping and GO functional categorization. Common gene IDs including microarray probe names can be automatically recognized. Pathway information was also compiled from databases like KEGG, Reactome, MSigDB, GSKB, and araPath. As an example, we analyzed an RNA-Seq dataset involving siRNA-mediated Hoxa1 knockdown in lung fibroblasts, and identified the down-regulation of cell-cycle genes, in agreement with previous studies. Our analyses also reveal the possible roles of E2F1 and its target genes, including microRNAs, in blocking G_1/S transition, and the upregulation of genes related to cytokines, lysosome, and neuronal parts. iDEP enables users to conduct in-depth bioinformatics analysis of transcriptomic data through a graphical interface. Freely available at http://ge-lab.org/idep/

P0103: Methods: Bioinformatics

RepeatExplorer Galaxy Server for In-Depth Characterization of Repetitive Sequences in Next-Generation Sequencing Data

Petr Novak, Pavel Neumann, Nina Hoštáková and Jiri Macas, Biology Centre CAS, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic

Repetitive DNA makes up large portions of plant and animal nuclear genomes, yet it remains the least-characterized genome component in most species studied so far. Recent availability of high-throughput sequencing data together with novel bioinformatics tools provide necessary resources for in-depth investigation of genomic repeats and enable large-scale repeat analysis to be run by biologically oriented researchers. Here we present RepeatExplorer Galaxy server (https://repeatexplorer-elixir.cerit-sc.cz), a collection of software tools for in-depth characterization of repetitive elements, which is accessible via web interface. A key component of the server is the computational pipeline using a graph-based sequence clustering algorithm to facilitate *de novo* repeat identification without the need for reference databases of known elements. Since its first release, number of tools for repeat analysis available on public server has grown up. Today RepeatExplorer include several tools which can be used on both short unassembled NGS reads and genome assemblies. New tools include automatic classification of repetitive sequences based on comprehensive database of transposable element protein domains, genome annotation and classification of repetitive elements, and improved identification of tandem repeats using TAREAN pipeline.

References:

Novak, P., Avila Robledillo, L., Koblizkova, A., Vrbova, I., Neumann, P., Macas, J. (2017) – TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 45(12): e111.

Novak, P., Neumann, P., Pech, J., Steinhaisl, J., Macas, J. (2013) – RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics 29: 792-793.

P0104: Methods: Bioinformatics

Dot: An Interactive Dot Plot Viewer for Comparative Genomics

Maria Nattestad[1], Michael C. Schatz[2] and Brett Hannigan[1], (1)DNAnexus, Mountain View, CA, (2)Johns Hopkins University, Baltimore, MD

Advances in long-read sequencing and scaffolding technologies are leading to unprecedented quality and quantity of genome assemblies. Comparing new assemblies to existing genomes of related species is crucial to understanding differences between organisms across the tree of life. The classic method for visualizing genome-genome alignments is the dot plot, which provides an excellent overview of alignments from the perspective of both genomes. However, dot plots have barely changed in the past decade and are still generated from the command-line as static images, limiting detailed investigation.

Here we present Dot, an interactive dot plot viewer that allows genome scientists to visualize genome-genome alignments in order to evaluate new assemblies and perform explorative comparative genomics. Dot enables scientists to explore regions of interest in detail by zooming in and inspecting unique and repetitive alignments. In addition to showing alignments, Dot allows scientists to load annotations for either or both genomes to show additional context, e.g. understanding how sequence differences map to gene differences. This might also allow scientists to explore how known repetitive elements in the reference genome affect assembly quality in specific regions. Dot supports the output of MUMmer, the most commonly used software method for aligning genome assemblies, with the potential to support outputs from future genome-genome alignment algorithms as they emerge. By leveraging D3 and canvas in JavaScript, Dot combines the benefits of interactivity with scalability, enabling scientists to explore large genomes and create publication-quality images. Dot is free, publicly available, and open source.

P0105: Methods: Bioinformatics

Bioinformatic Analysis using Jetstream, a Cloud Computing Environment

Bhavya Papudeshi[1], Sheri A. Sanders[2], Carrie Ganote[1], Jeremy Fischer[1] and Tom Doak[1], (1)Indiana University, Bloomington, IN, (2)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN

National Center for Genome Analysis Support (NCGAS) assists researchers in addressing the scientific challenges of understanding and analyzing the wealth of gene sequence information now available. This includes on-boarding biology professionals who lack the necessary computational background to run their analyses on high-performance computing systems. Virtual machines help with the transition to command line use, software installation, and running analysis in the Linux environment (as most high-performance computing clusters). Jetstream (https://jetstream-cloud.org/) is a cloud computing resource that provides access to preconfigured virtual machines, making the transition relatively effortless, flattening the learning curve needed to get results from experiments that otherwise produce an untenable amount

of data. Currently, over 14% of all allocations of usage on Jetstream are for biology other than protein folding – the majority of this being some sort of genomic analysis. NCGAS currently hosts over 123 genome analysis and bioinformatics software titles on Jetstream as preconfigured virtual machines. In this digital tool and resources workshop, we will demonstrate on how to set up Jetstream accounts, start a preconfigured virtual machine, and run genomic analysis on this virtual machine (https://ncgas.org/Blog_Posts/Getting%20Started%20on%20Jetstream.php). Jetstream also provides environments for prototyping and publishing tailored workflows that gives researchers access to interactive computing and data analysis resources on demand.

P0106: Methods: Bioinformatics

XMView: A Multiple Alignment XMap Viewer with Genetic Map Integration

Steve Wanamaker, Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA

XMView is a program for drawing complete clusters of Refaligner optical molecules and contigs. With this program it is possible to display two optical maps at the same time, enabling comparisons between the optical maps. A track of genetic map coordinates allows the user to see the linearity of the assembly. Both of these features help the user to discover chimeras in contigs and optical molecules.

P0107: Methods: Bioinformatics

Gene Mapping by Segregation: From Mendel to BSA, and Beyond

Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA

The observations on segregation of pea flowers by George Mendel led to the discoveries of the first two inheritance laws: segregation and independent assortment. If the connectional mapping from phenotypes to genetics is considered as the first-generation of gene mapping, the second generation gene mapping should be the Morgan law that the degree between complete co-segregation and independent assortment is proportional to genetic distance. The Morgan law has been widely used to map major genes through pedigree segregant analysis and Quantitative Trait Loci (QTLs) for complex traits. The major challenge of QTL mapping is the resolution caped the recombination events occurred during recent generation. The reverse strategy was to use the Linkage Disequilibrium (LD) that could be remained after many generation of random matting if the genetic loci are close enough. This third-generation gene mapping, Genome-Wide Association Study (GWAS), plays the major role even today. To exclude the possibility of LD due to non-physically linked reason, homogenous populations, such as F2, were suggested as alternatives. To gain enough meiosis in the homogenous populations, large number of individuals are required. Therefore, the fourth-generation gene mapping pools individuals at with extreme phenotype distribution for deep sequencing to reveal the allele frequencies. This Bulked Segregant Analysis (BSA) gain both on false positive control and mapping resolution at the cost of making crosses. Based on reviewing the first four generation gene mapping methods, the prospective properties are discussed for the next generation gene mapping.

P0108: Methods: Bioinformatics

BRAKER2: Incorporating Protein Homology Information into Gene Prediction with GeneMark-EP and AUGUSTUS

Katharina Hoff[1], Alexandre Lomsadze[2], Mario Stanke[1] and Mark Borodovsky[3], (1)University of Greifswald, Greifswald, Germany, (2)Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering,Georgia Institute of Technology, Atlanta, GA, (3)Joint Georgia Tech and Emory Wallace H Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA

The rapidly growing number of genome sequencing projects requires fully automated methods for accurate gene structure annotation. To this end, we previously developed BRAKER1: In presence of RNA-Seq data, BRAKER1 allows fully automated gene structure annotation in novel genomes with GeneMark-ET and AUGUSTUS. The novel BRAKER2 additionally allows for the integration of protein homology information. BRAKER2 fully automatically constructs training genes either on the basis of RNA-Seq or protein sequence data. If RNA-Seq data is input, training of GeneMark-ET and AUGUSTUS is performed on the basis of these data; otherwise, GeneMark-EP and AUGUSTUS are trained on the basis of homologous protein sequence data. Final predictions incorporate evidence from all input sources.
BRAKER2 is available for download at
http://bioinf.uni-greifswald.de/bioinf/braker/ and http://exon.gatech.edu.

P0109: Methods: Bioinformatics

Combining RNA-Seq Data and Homology-Based Gene Prediction for Plants, Animals and Fungi

Jens Keilwagen[1], Frank Hartung[1], Michael Paulini[2], Sven O. Twardziok[3] and Jan Grau[4], (1)Julius Kühn-Institut, Quedlinburg, Germany, (2)European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom, (3)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (4)Institute of Computer Science, Martin Luther University Halle–Wittenberg, Halle, Germany

Genome annotation is of key importance in many research questions. The identification of protein-coding genes is often based on transcriptome sequencing data, ab-initio or homology-based prediction. Recently, it was demonstrated that intron position conservation improves homology-based gene prediction, and that experimental data improves ab-initio gene prediction.
Here, we present an extension of the gene prediction tool GeMoMa that utilizes amino acid sequence conservation, intron position conservation and optionally RNA-seq data for homology-based gene prediction. We show on published benchmark data for plants, animals and fungi that GeMoMa performs better than the gene prediction programs BRAKER1, MAKER2, and CodingQuarry, and purely RNA-seq-based pipelines for transcript identification. In addition, we demonstrate that using multiple reference organisms may help to further improve the performance of GeMoMa. Finally, we apply GeMoMa to four nematode species and to the recently published barley reference genome indicating that current annotations of protein-coding genes may be refined using GeMoMa predictions.
GeMoMa might be of great utility for annotating newly sequenced genomes but also for finding homologs of a specific gene or gene family. GeMoMa has been published under GNU GPL3 and is freely available at http://www.jstacs.de/index.php/GeMoMa.

P0110: Methods: Bioinformatics

COPO - a Web Platform for "FAIR" Data in Plant Science

Felix Shaw, Earlham Institute, Norwich, United Kingdom

COPO is a portal for plant scientists to describe, store and retrieve data more easily. Data description is critical to increase the value of the data itself, allowing scientists (and online search tools) to better understand its relevance.

COPO assists scientists with labelling and tagging their work, in other words 'contextualising research', so that it is found at the right time and place. It seeks to make it easy for scientists to share their results, by making helpful suggestions for what information you might want to submit, based on past submissions and similar workflows from other scientists. It will also provide a way to get credit for work, such as source code or research data which don't typically get citations in scientific papers, even though they are crucial for actually undertaking research. Publications, data, images, and other 'research objects' can be submitted through COPO to remote long-term storage repositories. By providing a helpful graphical user interface (GUI), COPO relieves much of the eye-watering formatting burden required for preparing the information that goes alongside the research data.

In technical terms, COPO creates, aggregates and describes research objects. It normalises metadata to specific controlled vocabularies (ontologies), assisting with integration and standardisation of data through consistent terminology, boosting reproducibility and impact of research.

P0111: Methods: Bioinformatics

LOGAN: A Framework for Lossless Graph-Based Analysis of High Throughput Sequence Data

Anthony Bolger, RWTH Aachen, Aachen, Germany, Alisandra Denton, Heinrich-Heine-Universität, Düsseldorf, Germany, Marie E Bolger, IBG-2, Forschungszentrum Jülich, Jülich, Germany and Bjoern Usadel, Forschungszentrum Jülich GmbH, Jülich, Germany

Recent massive growth in the production of sequencing data necessitates matching improvements in bioinformatics tools to effectively utilize it. Existing tools suffer from limitations in both scalability and applicability which are inherent to their underlying algorithms and data structures. We identify the key requirements for the ideal data structure for sequence analyses: it should be informationally lossless, locally updatable, and memory efficient; requirements which are not met by data structures underlying the major assembly strategies Overlap Layout Consensus and De Bruijn Graphs. We therefore propose a new data structure, the LOGAN graph, which is based on a memory efficient Sparse De Bruijn Graph with routing information. Innovations in storing routing information and careful implementation allow sequence datasets for Escherichia coli (4.6Mbp, 117x coverage), Arabidopsis thaliana (135Mbp, 17.5x coverage) and Solanum pennellii (1.2Gbp, 47x coverage) to be loaded into memory on a desktop computer in seconds, minutes, and hours respectively. Memory consumption is competitive with state of the art alternatives, while losslessly representing the reads in an indexed and updatable form. Both Second and Third Generation Sequencing reads are supported. Thus, the LOGAN graph is positioned to be the backbone for major breakthroughs in sequence analysis such as integrated hybrid assembly, assembly of exceptionally large and repetitive genomes, as well as assembly and representation of pan-genomes.

P0112: Methods: Bioinformatics

SciApps: A Web-Based Platform for Reproducible Bioinformatics Workflows

Zhenyuan Lu, Liya Wang, Peter Van Buren and Doreen Ware, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

There are increasing needs to store and analyze data on distributed storage and computing systems imposed by the rapid growth of both sequence and phenotype data generated by high-throughput methods. A workflow management system is needed to ensure efficient data management across heterogeneous systems, simplify the task of analysis through automation, and make large scale bioinformatics analysis accessible and reproducible.

To address these needs, we have developed SciApps, a web-based platform for reproducible bioinformatics workflows. The system is fully integrated with CyVerse Cyber-infrastructure (CI) through the Agave platform for job management and iRODS-based CyVerse Data Store for data management. To create a workflow, each analysis job is submitted, recorded, and accessed through the web portal. Part or all of a series of recorded jobs can be saved as reproducible, sharable workflows for future execution using the original or modified inputs and parameters. The platform is designed to automate the execution of modular Agave apps and make it easy to bring reproducible workflows to cloud-based computing systems.

P0113: Methods: Bioinformatics

Incentivizing Volunteer Curation through Micropublication

Paul W. Sternberg, California Institute of Technology, Pasadena, CA

WormBase is the authoritative database of genomic and genetic data of the free living nematode, *Caenorhabditis elegans,* a major model organism for biomedical research. WormBase houses the entire genome for this nematode and related nematodes including many parasites and agricultural pests. The gene models of all of these species are continually updated with an emphasis on those species most intensively used. In addition, WormBase curators collect and annotate gene function data from the published primary literature. As a result, we provide user-friendly report pages for over 39 data classes, encompassing nearly 12 million unique pages. Curating and integrating biological and biomedical knowledge into computable publicly available resources is a key step to expedite new discoveries but is costly and time consuming. So-called community annotation in which researchers volunteer effort to associate metadata with data or integrate information has proven difficult to incentivize. We have established a new data capture and dissemination paradigm that automatically and simultaneously captures and ingests biomedical data into our repository and publishes them in an online, peer-reviewed, open access journal *'Micropublication: biology'*. This new platform introduces a curation paradigm shift, allowing authors to directly submit the output of their research into the database using pre-designed intelligent web forms. Simultaneously, the process will automatically generate a 'publication-like' PDF file that will be publishable and citable according to findable, accessible, interoperable and reproducible (FAIR) data principles. We call these single result experiments, streamlined with minimal narrative, "micropublications." Authors preserve provenance and establish credit for their research and the automated flow of data they submit will be made publicly accessible in WormBase, integrated with existing datasets that have been

manually extracted from the literature for almost 2 decades. In addition, researchers will be able to share both positive and negative data with the scientific community, fulfilling funding agencies' requirements to share all data coming from publicly funded research for further re-use. While we, along with most information resources, have spent effort on eliciting curatorial action from our communities, the Micropublication:biology platform provides an added incentive of a bona-fide citation that will eventually be indexed and findable in major citation indexers such as PubMed. After establishing this data retrieval/publication pipeline with WormBase first, and the other Model Organism Database (MOD) members of the Allied Genome Resources (AGR): FlyBase, Mouse Genome Database (MGI), Rat Genome Database (RGD), Saccharomyces Genome Database (SGD), Zebrafish Model Organism Database (ZFIN), we will work to expand to non-member, but otherwise critical organismal databases, such as Xenbase (*Xenopus laevis* and *tropicalis* Database), DictyBase (*Dictyostelium discoideum* database), PomBase (*Schizosaccharomyces pombe* Database), among others. I will discuss our progress and potential broader use of this incentivization approach.

P0114: Methods: Bioinformatics

A Practical Haplotype Graph for Determining Genomic Sequence

Lynn C Johnson[1], Peter Bradbury[2], Terry Casstevens[1], Daniel Ilut[3], Zachary Miller[1], Ramu Punna[1], M. Cinta Romay[1] and Edward Buckler[4], (1)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (2)Cornell University/USDA-ARS, Ithaca, NY, (3)Cornell University, Ithaca, NY, (4)USDA-ARS, Cornell University, Ithaca, NY

Improved sequencing technologies and declining genotyping costs facilitate the acquisition of large genomic datasets. The Practical Haplotype Graph (PHG) is a computational framework aimed at processing these datasets to infer high-density genotypes from low-coverage sequence. The PHG framework combines existing genomic analysis software with custom code Docker images for deployment at user sites. Data from high depth WGS reads, genome assemblies and a reference genome are loaded to a relational database with reference genome coordinates attached to anchor nodes representing conserved regions. Short reads from skim sequences are then aligned against database consensus sequences to obtain haplotype counts. These are combined with pedigree information and used with a hidden Markov model to find the most likely path through the Haplotype Graph. The resulting path can be translated to variant calls and output in VCF format.
By integrating an entire species worth of prior information, the PHG pipeline can produce an accurate whole genome sequence from any sequencing approach. This has applications from basic genomic research into chromatin structure to applied plant breeding.

P0115: Methods: Bioinformatics

LTR_retriever: A Highly Accurate and Sensitive Program for Identification of LTR Retrotransposons

Shujun Ou and Ning Jiang, Michigan State University, East Lansing, MI

Long terminal-repeat retrotransposons (LTR-RTs) are prevalent in plant genomes. Identification of LTR-RTs is critical for achieving high-quality gene annotation. Based on the well-conserved structure, multiple programs were developed for *de novo* identification of LTR-RTs; however, these programs are associated with low specificity and high false discovery rate (FDR). Here we report LTR_retriever, a multithreading empowered Perl program that identifies LTR-RTs and generates high-quality LTR libraries from genomic sequences. LTR_retriever demonstrated significant improvements by achieving high levels of sensitivity (91.8%), specificity (94.7%), accuracy (94.3%), and precision (90.6%) in model plants. LTR_retriever is also compatible with long sequencing reads. With 40k self-corrected PacBio reads equivalent to 4.5X genome coverage in Arabidopsis, the constructed LTR library showed excellent sensitivity and specificity. In addition to canonical LTR-RTs with 5'-TG…CA-3' termini, LTR_retriever also identifies non-canonical LTR-RTs (non-TGCA), which have been largely ignored in genome-wide studies. We identified seven types of non-canonical LTRs from 42 out of 50 plant genomes. The majority of non-canonical LTRs are *Copia* elements, with which the LTR is four times shorter than that of other *Copia* elements, which may be a result of their target specificity. Strikingly, non-TGCA *Copia* elements are often located in genic regions and preferentially insert nearby or within genes, indicating their impact on the evolution of genes and potential as mutagenesis tools.

P0117: Methods: Cellular Processes and Regulatory Networks

Dissecting SDG Lignan Biosynthetic Pathways in EMS Mutagenized Flax using Ampliseq Gene Panels

Bourlaye Fofana[1], Kaushik Ghose[2], Ashok Somalraju[1], Jason McCallum[1] and Sylvie Cloutier[3], (1)Agriculture and Agri-Food Canada, Charlottetown, PE, Canada, (2)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada

Ethyl methane sulfonate (EMS) mutagenesis is now a common practice in plant breeding. Whereas forward genetic helps identifying the desired phenotypic traits, genotyping the allelic variants responsible for the phenotypic variation is always challenging due to the random nature of the induced mutations along the pathways. In the absence of full genome sequence, Ampliseq gene panel has been a good option for dissecting biosynthetic pathways. Here, an Ampliseq gene panel consisting of 11 genes was used to dissect and ascertain EMS mutations in the SDG lignan biosynthetic pathway from 84 M4 flax plants. The data not only confirmed previous amplicon sequencing and KASP genotyping results in *UGT74S1* gene, but also contributed to identifying additional mutations, albeit sense or of minor effects, in other genes of the pathway. The usefulness of Ampliseq gene panel in pathways analysis will be discussed.

P0118: Methods: Cellular Processes and Regulatory Networks

RNAseq Transcriptomes Deciphering Differential Gene Expression between Resistant and Susceptible Potato Cultivars to Common Scab (*Streptomyces scabies*)

Bourlaye Fofana, Ashok Somalraju and David Main, Agriculture and Agri-Food Canada, Charlottetown, PE, Canada

Potato scab (*Streptomyces scabies*) is widespread in all potato growing regions and constitutes a serious constraint for potato marketability. Scab management relies mainly on agronomic practices and breeding for resistant cultivars. Thus, identification of key genes and pathways contributing plant defense system and signaling in response to the disease appears to be a pre-requisite toward developing genome editing strategies for scab resistance in potatoes. Significant differences were observed between the cultivars for disease responses and for

differentially expressed genes (DEGs). The data will be presented and discussed in light of GO and KEGG pathway enrichment towards deciphering the signaling mechanisms of scab resistance in potato.

P0119: Methods: Cellular Processes and Regulatory Networks

Organellomics of Stress Response in Plants: Phenotyping Peroxisome Abundance

Deirdre Fahy, Kathleen Hickey and **Andrei Smertenko**, Institute of Biological Chemistry, Washington State University, Pullman, WA

Plant peroxisomes maintain a plethora of key life processes including fatty acid β-oxidation, photorespiration, synthesis of hormones, and reactive oxygen species (ROS) homeostasis. The abundance of peroxisomes in cells can change in response to environmental cues, however the significance of this phenomenon remains unknown. The progress in this direction is hindered by the lack of an efficient method for measuring peroxisome abundance in plant tissues. Counting peroxisomes using fluorescence or electron microscopy is expensive and time-consuming. Here we report development of a high-throughput technique for measuring peroxisome abundance using the small fluorescent probe Nitro-BODIPY. We applied this technique to analyze peroxisome abundance during plant development, salt, and drought stress in *Arabidopsis thaliana* and *Triticum aestivum*. We found that salt stress promotes peroxisome proliferation in tolerant genotypes and inhibits peroxisome proliferation in the susceptible genotypes. Peroxisome abundance was also higher in drought-stress plants. Principal component analysis shows negative correlation between yield parameters and peroxisome abundance under drought. In conclusion, peroxisome abundance can be used as a proxy for stress-tolerance.

P0120: Methods: Cytology

Sequencing of Single Pollen Nuclei Reveals Meiotic Recombination Events at Megabase Resolution

Steven Dreissig[1], Joerg Fuchs[2], Axel Himmelbach[3], Martin Mascher[4] and **Andreas Houben**[2], (1)IPK, Gatersleben, Germany, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, (3)Institute of Plant Genetics & Crop Plant Research, Gatersleben, Germany, (4)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany

Meiotic recombination is a fundamental mechanism to generate novel allelic combinations which can be harnessed by breeders to achieve crop improvement. The recombination landscape of many crop species, including the major crop barley, is characterized by a dearth of recombination in 65% of the genome. In addition, segregation distortion caused by selection on genetically linked loci is a frequent and undesirable phenomenon in double haploid populations which hampers genetic mapping and breeding. Here, we present an approach to directly investigate recombination at the DNA sequence level by combining flow-sorting of haploid pollen nuclei of barley with single-cell genome sequencing. We confirm the skewed distribution of recombination events towards distal chromosomal regions at megabase resolution and show that segregation distortion is almost absent if directly measured in pollen. Furthermore, we show a bimodal distribution of inter-crossover distances, which supports the existence of two classes of crossovers which are sensitive or insensitive to physical interference. We conclude that single pollen nuclei sequencing is an approach capable of revealing recombination patterns in the absence of segregation distortion.

P0121: Methods: Cytology

Cytogenetics of Duckweed

Phuong Hoang, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben,, Germany

Neotenous aquatic duckweeds belong to monocot order Alismatales and comprise 37 species of 5 genera: Spirodela, Landoltia, Lemna, Wolffiella and Wolffia. The genome size and chromosome number of duckweed species vary from 158 Mbp *(Spirodela polyrhiza)* to 1881 Mbp *(Wolffia arrhiza),* and from 2n = 20 to 126, respectively. We focused on Spirodela - the ancestral genus: (1) to establish a reference genome map for *S. polyrhiza* and (2) to study chromosome rearrangements between the only two species *S. polyrhiza* and *S. intermedia,* both with similar genome size (~158 Mbp).

We applied comparative FISH with BACs and pooled BAC probes on seven *S. polyrhiza* clones to address the reason of discrepancies between previous maps of *S. polyrhiza.* Our result revealed no chromosome rearrangements between the seven studied clones and integrated Oxford Nanopore sequencing data to establish an updated reference genome map for *S. polyrhiza*.

The same approach was used to investigate the chromosome homeology and karyotype evolution between *S. polyrhiza* (n=20) and *S. intermedia* (n=18). Two scenarios of karyotype evolution are supposed, considering the ancestral karyotype was similar either to that of *S. polyrhiza* or that *S. intermedia*.

Key words: *S. polyrhiza, S. intermedia,* cytogenetic map, chromosome rearrangements, karyotype evolution.

P0122: Methods: Functional Analysis

RefEx, a Reference Gene Expression Dataset As a Web Tool for the Functional Analysis of Genes

Hiromasa Ono and Hidemasa Bono, Database Center for Life Science, Mishima, Japan

RefEx (Reference Expression dataset; http://refex.dbcls.jp) is a web tool for browsing reference gene expression, which provides access to curated data from several other public databases, with expression levels in forty tissues measured by four well-established gene-expression quantification technologies.

"RefEx" contains the data from three kinds of organisms, human, mouse, and rat, obtained from normal tissues and cell lines (556 tissues/cell lines in total) measured by four different methods(EST, GeneChip, CAGE, RNA-seq), All data are acquired from public database including those from FANTOM5 project.

Along with an extensive collection of gene expression data above "RefEx" enables the comparison of the gene expression status in each tissue/cell with the difference among the measurement methods. You can search the data simply by gene name, or gene ontology and family name to obtain the data for certain group of genes. Furthermore, you can select "tissue/cell-specific genes", namely marker genes representing

characteristics of the tissue/cell calculated by applying a uniform method to all the accumulated public data by clicking tissue icons in "RefEx" top page.
The search/select result shows the comparison of the relative expression levels among tissues and among the four measurement methods, and the relative expression amount is reflected as a heat-map in the 3D model of the human body. You can also compare annotation information (Gene Ontology, etc.) on functions assigned to genes in the search results. These functions support new knowledge discoveries and hypothesis buildings.
Using "RefEx", researchers can confirm the expression level of the genes of interest in many tissues or cells under normal condition, without bench-top experiments. It is also useful as a tool to know the relationship of genes found in functional analysis to elucidate biological phenomena and interpret research results leading to the development of medicines etc. Thus "RefEx" is expected to contribute a wide variety of life science research as a powerful web tool for gene expression analysis.
The RefEx paper published in Scientific Data.

P0123: Methods: Functional Analysis
Expanding the Critical Assessment of Function Annotation with Experimental Data and Biocuration
Naihui Zhou[1], Yuxiang Jiang[2], Timothy Bergquist[3], Alex W. Crocker[4], Kimberley A. Lewis[4], Deborah A. Hogan[5], Maria J Martin[6], Claire O'Donovan[7], Casey Greene[8], Sean D. Mooney[3], Predrag Radivojac[2] and **Iddo Friedberg**[1], (1)Iowa State University, Ames, IA, (2)Indiana University, Bloomington, IN, (3)University of Washington, Seattle, WA, (4)Dartmouth College, Hanover, NH, (5)Dartmouth Coollege, Hanover, NH, (6)European Bioinformatics Insitute, Hinxton, United Kingdom, (7)European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, (8)University of Pennsylvania, Philadelphia, PA
The increasing volume and variety of genotypic and phenotypic data is a major defining characteristic of modern biomedical sciences. Assigning functions to biological macromolecules, especially proteins, turned out to be one of the major challenges to understand life on a molecular level. While molecular experiments provide the most reliable annotation of proteins, their relatively low throughput and restricted purview have led to an increasing role for computational function prediction. However, properly assessing methods for protein function prediction and tracking progress in the field remain challenging as well. The Critical Assessment of Functional Annotation (CAFA) is a timed challenge to assess computational methods that automatically assign protein function. Here we report the preliminary results of the third CAFA challenge, and outline some additions that have taken place since the second CAFA. One hundred and forty seven models from 66 research groups were received and are currently being evaluated for accuracy in predicting protein function in 27 target species. These functions are described by the Gene Ontology (GO) and the Human Phenotype Ontology (HPO). Comparisons between top-performing methods in CAFA1 and CAFA2 showed significant improvement in prediction accuracy, demonstrating the general improvement of automatic protein function prediction algorithms. We expect to see more improvement in CAFA3. CAFA3 features expanded protein sets for predictions. We are using experimental whole genome screens to generate ground truths for select functions in *Drosophila melanogaster*, *Pseudomonas aeruginosa,* and *Candida albicans*. Additionally, we released sets of moonlighting proteins, to further challenge function prediction methods.

P0124: Methods: Functional Analysis
Ultrafast Analyses of NGS Datasets: Cleaning & Differential Expression
Aurelie Kapusta[1,2], Steven Flygare[3], Chase Miller[1,2], Yi Qiao[1,2], Gabor Marth[1,2], Edgar J. Hernandez[1,2], Guochun Liao[3], Martin G. Reese[3], Robert Schlaberg[4,5] and Mark Yandell[1,2], (1)Department of Human Genetics, University of Utah, Salt Lake City, UT, (2)USTAR Center for Genetic Discovery, Salt Lake City, UT, (3)IDbyDNA Inc., San Francisco, CA, (4)Department of Pathology, University of Utah, Salt Lake City, UT, (5)ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT
NGS-based metagenomics has the potential to revolutionize science, but computation times have hindered adoption. In response, researchers at the University of Utah, the Centers for Disease Control and Prevention (CDC), ARUP Laboratories, and IDbyDNA Inc. have developed Taxonomer [1,2] – an ultrafast engine for comprehensive metagenomics data analysis and interactive results visualization (https://www.taxonomer.com). Taxonomer is unique in providing integrated nucleotide and protein-based classification, and it is extremely fast: Taxonomer can search every read in an Illumina RNA-seq dataset of ~7 million reads against a database of ~38 million sequences, identifying the most likely organism of origin for every read in less than 10 minutes using 16 threads [1]. Taxonomer thus opens new avenues for research applications. We will present two broad ones: RNA transcript profiling and detection of NGS sample contamination. For the latter, we will use examples from human and other organisms to demonstrate how researchers can use Taxonomer to profile metagenomics dataset; identify contamination in NGS datasets; and profile transcript abundances. We will highlight the step by step tutorial for building custom databases which enable researchers to customize Taxonomer for their own organisms of interest, specialized applications, and to embed Taxonomer in their bioinformatics workflows. NGS has already revolutionized genetics; now ultrafast metagenomics is about to revolutionize comparative genomics and our results provide a preview.
[1] Flygare, Simmon et al. (2016). *Genome Biology*
[2] https://github.com/Yandell-Lab/taxonomer_0.5 (Copyright (c) 2016 IDbyDNA Inc.)

P0125: Methods: Functional Analysis
PE: *Phenotype Substitute for Environment*, Proposing a New Parameter for G×E Study
Takanari Tanabata[1], Sachiko Isobe[2], Atsushi Hayashi[1], Hidenori Tanaka[3], Masatsugu Hashiguchi[3], Akihiro Nakaya[4], Masataka Kikuchi[5], Mai Hasegawa[5], Shusei Sato[6], Sayuri Tanabata[7] and Ryo Akashi[3], (1)Kazusa DNA Research Institute, Chiba, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)University of Miyazaki, Miyazaki, Japan, (4)Graduate School of Medicine, Osaka University, Suita, Osaka, Japan, (5)Osaka University, Suita, Osaka, Japan, (6)Tohoku University, Sendai, Japan, (7)Ibaraki University, Ibaraki, Japan
Consideration of genotype × environment interaction (G×E) is becoming more important in genetic studies such as genome wide association analysis (GWAS) and genomic selection (GS). Next generation sequencing technologies enable obtaining whole genome sequencing data in

each individual of an analyzed population. However, getting comprehensive environmental data affecting to each plant is almost impossible, because environment condition surrounding a plant is diverse in terms of types of environmental factors and scale (micro to macro level). The imbalance quality and quantity of data between genome and environmental values makes incomplete G×E studies. Hence, we propose a new parameter, PE (Phenotype substitute for Environment), for more accurate G×E studies. PE is a parameter generated from phenotypic values of each plant during a growing period, obtaining by traditional and/or digital measurement methods such as image analysis. It substitutes environmental factors that cannot measured by sensors set in the test space. The concept of PE is generated from a fact that a phenotypic value is able to be explained by comprehensive environmental factors during the growing stage. To realize the concept of PE, we are obtaining phenotypic values in soybean and *Lotus japonicus* across Japan by using digital image analysis. The phenotypic values are dissected in phytomer unit so that a factor, time, is reflected to the G×E studies. This work is supported by JST CREST Grant Number JPMJCR16O1, Japan.

P0126: Methods: Functional Analysis

Non-Ripening Is Required for Efficient Agrobacterium Infection

Haoting Chen[1], **Ning WANG**[2], Satoko Nonaka[1], Miyako Kusano[1] and Hiroshi Ezura[1], (1)Graduate School of Life and Environmental Sciences, University of Tsukuba, Japan, (2)Graduate school of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Plant growth regulator ethylene is apparently related to flower formation, fruits growth and ripening. Whereas ethylene is one of the factors that inhibits *Agrobacterium*-mediated gene transformation. Tomato *Non-ripening* (*Nor*) mutant represses most phenotypic signs of fruit ripening because it fails to produce autocatalytic ethylene. Although high efficient protocol for genetic transformation in Micro-Tom has been established, *nor* mutant revealed lower frequency of gene transfer in compared with WT. The infection efficiency could not simply be improved through adjusting the concentration of plant hormone, auxin and cytokinin, in culture medium. Our previous study demonstrated that knockout mutant of *Nor* in impaired for ripening related genes particularly inhibits genes expression of *ACO* and transcription factor *RIN*, and ethylene-insensitive mutants were expected to enhance the bacterial growth. Nevertheless, *nor* mutant showed significant higher ethylene evolution rate thereafter infection by *Agrobacterium*. To suppress ethylene level in host plant during cocultivation, the 1-aminocyclopropane-1-carboxylate (ACC) deaminase were carried out, which cleaves the immediate ethylene precursor ACC. We infected plants with *Agrobacterium* maintaining multiple vector construct, vector for ACC deaminase gene together with vector for β-glucuronidase (GUS) reporter gene. Infection efficiency and expression level of ethylene synthesis related genes were further analyzed. Our observations indicated suppression of ethylene production lead to high efficiency of agrobacterium-mediated genetic transformation in *nor* mutant. Evidences suggested alternative regulatory mechanisms of *Nor* underlying ethylene synthesis between fruits ripening and wounding.

P0127: Methods: Functional Analysis

Functional and Comparative Genomics of the Water-Use Efficient Crassulacean Acid Metabolism Adaptation of Photosynthetic CO2 Fixation in the Genus Kalanchoë

Susanna F. Boxall[1], Louisa V. Dever[1], Nirja Kadu[1], Nina Pugh[1], Yongxiang Fang[1], Rachel Brenchley[2], Neil Hall[3], Alistair Darby[1] and **James Hartwell**[1], (1)University of Liverpool, Liverpool, United Kingdom, (2)University of Salford, Salford, United Kingdom, (3)Earlham Institute, Norwich, United Kingdom

Forward engineering of Crassulacean acid metabolism (CAM) into non-CAM crop species in order to enhance their water use efficiency requires a comprehensive knowledge of the minimal 'parts-list' for CAM. Here we describe progress with functional genomics research that aims to define and characterize the complete CAM genetic blueprint from the model species *Kalanchoë fedtschenkoi* and *Kalanchoë laxiflora*. A draft assembly of the 246 Mbp diploid *K. fedtschenkoi* genome has recently been further improved through the addition of 100X coverage of PacBio long reads. Quantitative RNA-seq analysis of light/ dark time course samples from C_3 and CAM leaves identified candidate CAM-associated genes. Many of these CAM-induced genes also undergo robust oscillations in transcript abundance over the light/ dark cycle, consistent with a role in the circadian optimization of the pathway. The RNA-seq data has guided the reconstruction of a comprehensive model of CAM wherein candidate CAM-recruited genes are allocated to each step in the biochemical pathway. This in turn has allowed targeted RNAi gene silencing and over-expression approaches to be applied to each candidate CAM gene through the generation of stable transgenic lines of *K. fedtschenkoi* and *K. laxiflora*. Detailed analysis of CAM-associated phenotypes in the transgenic lines is revealing which genes are essential for efficient CAM, and which genes are dispensable. Data will be presented for RNAi lines of *K. laxiflora* lacking key CAM genes. Several lines fail to fix atmospheric CO_2 in the dark period, and the phenotypic consequences of this will be described.

P0128: Methods: Functional Analysis

Identification of Genes Controlling Root System Architecture in Rice Using a Gel-Based Imaging System

Kevin Lehner, Benfey Lab; Duke University Program in Genetics and Genomics, Durham, NC

An uncertain climate, paired with rapid human population growth, presents a major challenge to maintaining food security in the twenty-first century. Improvement of cultivation of rice, a primary source of calories for nearly half of the world's people, provides a unique opportunity to address this challenge. As breeding in cereals has largely focused on aboveground phenotypes, root system traits represent potential unexplored targets for stress resilience and yield improvement. However, our understanding of the genetic control of root system architecture (RSA) in rice is fundamentally insufficient to contribute to these goals. The identification of novel rice RSA loci and the genes that underlie them could potentially provide breeders the tools to test the effect of root traits on water and nutrient usage.

Our lab has developed a gel-based imaging and phenotyping system to facilitate genetic mapping of root traits in rice. We have identified multiple root quantitative trait loci (QTL) in a biparental (Azucena x Bala) mapping population. Many of these QTL colocalize in distinct clusters in the genome, potentially representing major RSA control loci. I am using our phenotyping platform to narrow one such cluster on chromosome 7 that likely influences the distribution of biomass throughout the root system. I am currently evaluating mutants for candidate genes within this cluster for root traits. The identification and functional characterization of the gene(s) underlying this QTL will expand our rudimentary understanding of the genetic control of RSA and provide potential targets to test in breeding programs.

P0129: Methods: Functional Analysis

Investigation of Antimicrobial Peptide Genes Associated with Fungus and Insect Resistance in Maize

Joseph Ali Noonan, Mississippi State University, Mississippi State, MS

Antimicrobial peptides (AMPs) are small defense proteins present in various organisms. Major groups of AMPs include beta-barrelin, hevein, knottin, lipid transfer protein (LTP), thionin, defensin, snakin, and cyclotide. Plant AMPs convey host plant resistance to pathogens such as fungi, viruses, and bacteria, with some of plant AMPs from the cyclotide family showing insecticidal function. In this research, a genome-wide investigation on maize antimicrobial peptide genes was conducted. Thirty-nine new maize AMPs were identified in addition to 7 known maize AMPs. Protein sequence analysis revealed 10 distinguishable maize AMP groups. Analysis of gene expression by qRT-PCR revealed 5 maize AMP genes significantly associated with insect or fungus resistance. Identification of maize antimicrobial peptide genes will facilitate the breeding of host plant resistance and improve maize production.

P0130: Methods: Functional Analysis

A Road Map Towards Enlightening the Potato Greening Mechanisms

Ashok Somalraju[1], **Bourlaye Fofana**[1], Kaushik Ghose[2] and David Main[1], (1)Agriculture and Agri-Food Canada, Charlottetown, PE, Canada, (2)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

Potato greening, known as sunburn, causes 2-3% loss at the farm gate and during postharvest storage. The phenomenon involves chlorophyll synthesis in the cortical parenchyma cells beneath the periderm and a simultaneous accumulation of toxic glycoalkaloid compounds in the light-exposed area. However, the genetic means toward its avoidance have been less explored and mechanisms of such genetic approaches are unknown. Potato tubers from ethyl methane sulfonate (EMS)-treated and non-treated control lines were exposed to light, and EMS lines tolerant to greening were identified. RNAseq transcriptomic study comparing a non-greening and a greening line was performed before light exposure and after six days of continuous light exposure. Significant differentially expressed genes (DEGs) were observed between controls and non-greening tubers. The data will be presented and discussed in light of GO and KEGG pathway enrichment toward deciphering mechanism of light induced-greening of potato tubers.

P0131: Methods: Functional Analysis

Optimization of GTF Files for Transcriptome Analysis in the Horse

Shavahn C. Loux[1], Pouya Dini[2], Theodore S. Kalbfleisch[3], Alejandro Esteller-Vico[1] and Barry A. Ball[1], (1)Department of Veterinary Science, University of Kentucky, Lexington, KY, (2)Department of Reproduction, Faculty of Veterinary Medicine, University of Gent, Merelbeke, Belgium, (3)University of Louisville, Louisville, KY

The accuracy of downstream RNA sequencing analysis is heavily dependent on the accuracy of the general transfer format (GTF) file, a file which catalogs annotation information about the gene structure and location. Without a complete, accurate GTF file it is difficult to ensure either accurate or precise quantification of genes. In well-annotated organisms such as the human or the mouse, both NCBI and ENSEMBL have accurate, complete annotation files readily available. However, the current annotation of the horse genome lacks the specificity and sensitivity necessary to fully assess gene expression. To optimize our data analysis, we wanted to ensure we were using the best annotation possible, therefore, we used the integrative genome viewer to objectively analyze the accuracy of six separate gtf files, including 3 published gtf files (NCBI, ENSEMBL, Mansour et al.) and three customized gtf files using different iterations of Cufflinks and Cuffmerge. Overall, the most accurate published gtf file was from Ensembl, as it was the most complete, albeit often truncated. The NCBI gtf file was accurate as well, but lacked a large number of transcripts. Ultimately, the most accurate and complete gtf file was created by running Cufflinks on each BAM file individually, then merging each resultant transcript.gtf file with Cuffmerge, using the Ensembl gtf file as a guide. This file contained the genes from the ENSEMBL annotation file where the truncation is minimized, and several novel transcripts identified.

P0132: Methods: High-throughput Methods

Impact of NGS Library Quantitation and Sizing Accuracy on Downstream Sequencing

Whitney A Pike, Kit-Sum Wong and Steve Siembieda, Advanced Analytical Technologies, Inc., Ankeny, IA

Quality control of next-generation sequencing (NGS) libraries is essential for obtaining good sequencing results. Capillary electrophoresis and microfluidics instruments have become increasingly important in determining the quality, quantity, and size of NGS samples due to the advantages of low sample input, high sensitivity and short analysis time compared to agarose gel electrophoresis.

The Fragment Analyzer™ Automated CE System (Advanced Analytical Technologies, Inc.) is a capillary electrophoresis instrument with a flexible throughput platform capable of analyzing 12, 48, or 96 samples simultaneously. It has been used for the reliable qualification and quantification of nucleic acids in a wide range of concentrations and sizes. When employed in an NGS pipeline, the Fragment Analyzer can assess all sample types throughout library construction, from the initial assessment of sample integrity through final qualification.

Here we compare the data quality from the Fragment Analyzer to a microfluidics instrument, examine their quantification and sizing accuracy of an NGS library, and investigate what effect this may have on sequencing results.

P0134: Methods: High-throughput Methods

Evolution of Regulatory Element Tissue-Specificity across Mammals and Birds

Morgan Wirthlin, Irene Kaplow, Alyssa Lawler, Easwaran Ramamurthy, Ashley Brown and **Andreas R Pfenning**, Carnegie Mellon University, Pittsburgh, PA

One of the fundamental questions in biology is how changes in nucleotides in the genome have evolved to encode the wide array of tissues and phenotypes found across mammals and birds. Despite the current abundance of genomic and epigenomic data, modeling how CRE activity relates to complex phenotypes has proven challenging due to the complexity of the regulatory code, poor conservation at the nucleotide level, and lack of a large-scale experimental system to test predictions *in vivo*.

Here, we will develop new computational and experimental techniques to trace the tissue-specificity of liver, brain cortex, and brain striatum CREs across the genomes of mammalian and avian species. First, we create a database of human CREs by combining publicly available H3K27ac with measures of open chromatin in liver and brain tissue. Next, we leverage multiple sequence alignments to map the regulatory elements across mammalian and avian genomes. A comparison of orthologous regions of the genome to model organism epigenomic experiments conducted in tissues shows strong conservation of CRE tissue specificity, even in species as distantly related as human and zebra finch. We find that convolutional neural network models trained in human can predict tissue-specific regulatory activity across distantly related species, suggesting that the regulatory code is also conserved. To validate candidate tissue-specific CREs in model organisms where primary tissues is unavailable, we have developed a global massively parallel reporter assay, where the regulatory activity of thousands of synthesized regulatory sequences are measured in serval mouse tissue of interest *in vivo*.

P0135: Methods: High-throughput Methods

Precision Phenotyping of Epicuticular Waxes Associated with Insect Resistance

Eduardo D. Munaiz and Michael J. Havey, USDA-ARS and University of Wisconsin, Madison, WI

Accurate phenotyping is imperative for linkage mapping and association genetics. Amounts and types of epicuticular waxes on the leaf surface are important for plant-insect interactions. In onion, specific wax profiles are associated with resistance to the insect pest *Thrips tabaci*. Epicuticular wax profiles and specific metabolites are determined using Gas Chromatography Mass Spectrometry (GCMS), and affected by genetic and environmental factors. To unravel environmental effects and best phenotyping approaches, we studied 15 onion accessions with a wide range of wax composition, including three double haploids (DHs) with copious amounts of waxes and a mutant line ("glossy") with low amounts of waxes. Accessions were evaluated for amounts and types of epicuticular waxes under greenhouse (GH) and field conditions. GCMS revealed that the ketone hentriocontanone-16 (H16) is the most prevalent wax on onion leaves. There was significant variation for H16 among accessions, and it was significantly higher on the waxy DHs and lowest on the mutant line. Amounts of H16 varied significantly across environments, even though relative quantities were consistent within environments. Although genotype by environment (GxE) interactions were not significant, GH conditions allowed for greater power of discrimination among accessions. This study supports biochemical analyses of epicuticular waxes for phenotyping and selection of plants for modified profiles associated with insect resistance.

P0136: Methods: High-throughput Methods

BSA-Seq : An Efficient Method to Decipher a Complex Trait on Poplar, a High Heterozygous Diploid Genome

Aurélie Canaguier[1], Véronique Jorge[2], Vanina Guérin[2], Odile Rogier[2], Vincent Segura[2], Aurélie Chauveau[1], Elodie Marquand[1], Aurélie Bérard[3], Marie-Christine Le Paslier[1], Catherine Bastien[2] and Patricia Faivre-Rampant[1], (1)INRA, US 1279 Etude du Polymorphisme des Génomes Végétaux, F-91000 Evry, France, EVRY, France, (2)INRA, UR0588 AGPF, Centre Val-de-Loire, Orléans, France, Orléans, France, (3)INRA, US1279 Etude du Polymorphisme des Génomes Végétaux, F-91000, Evry, France

The efficiency of the Bulk Segregant Analysis (BSA) had clearly been demonstrated to detect genomic regions and genes involved in diverse traits. It allows large experiments reducing the cost and time and preserving the power of full individual's population analysis. These past few years the combination of BSA and New Generation Sequence (NGS) data (BSA-Seq) gave a new accuracy and depth to the discovery on many traits of interest, mainly on crop and model species.

In our study, we applied the BSA-Seq in a heterozygous and diploid genome context. We worked on the progenies derived from an interspecific cross *Populus deltoides* x *Populus trichocarpa* in which segregates the resistance to *Melampsora larici populina* (*Mlp*) leaf rust. We detected DNA variations with freebayes/0.9.21 and the soft masked genome of *Populus trichocarpa* Nisqually v3.0 as reference. Analysing DNA variations in between parents and bulks we obtained 27 regions or Quantitative Trait Loci based on NGS analysis (QTL-Seq) which could explain the resistance to *Mlp*. We first evaluated the strategy retrieving a previously cloned *Mlp* resistance gene governing the uredinia size in *Populus trichocarpa* clone 101-74 (R_{US}). Then we identified genomic markers which should better characterize this locus.

We demonstrated that, in our context, BSA-Seq allows to improve the fine mapping of a major gene. So, it can be a promising method on a high heterozygous diploid genome as Poplar to decipher complex trait. Next step is to proceed with it to fine map the other QTLs.

P0137: Methods: High-throughput Methods

Massive Phenotyping of Multiple Cranberry Populations Reveals Novel QTLs for Fruit Anthocyanin Content and Other Important Chemical Traits

Luis A Diaz Garcia[1,2], Brandon Schlautman[3], Giovanny Covarrubias-Pazaran[1], Andrew F. Maule[1], Nicholi Vorsa[4], James Polashock[5] and Juan E. Zalapa[6], (1)University of Wisconsin-Madison, Madison, WI, (2)Instituto Nacional de Investigaciones Forestales y Agricolas, Pabellon de Arteaga, Mexico, (3)The Land Institute, Salina, KS, (4)Rutgers University, New Brunswick, NJ, (5)USDA-ARS, Genetic Improvement of Fruits and Vegetables Laboratory, Chatsworth, NJ, (6)USDA, ARS, Madison, WI

Because of its benefits for human health, American cranberry (*Vaccinium macrocarpon* L.) production and commercialization around the world have gained importance in recent years. The presence of flavonoid compounds as well as the balance of sugars and acids are key quality characteristics of fresh and processed cranberry products. In this paper, we report the presence of novel QTL that affect the total anthocyanin content (TAcy), tritatable acidity (TA), proanthocyanidin content (PAC), Brix and mean fruit weight (MFW) in cranberry fruits. In total, more than 50 QTL were discovered in at least two of the three years evaluated in this study, with the most QTL identified for the TAcy and MFW traits. In addition, we demonstrate the utility of digital imaging as a reliable, inexpensive and high-throughput strategy for the quantification of anthocyanin content in cranberry fruits. Using this imaging approach, we identified a set of QTL across three different breeding populations which collocated with the anthocyanin QTL identified using a wet chemistry approach. Finally, we discuss the presence of candidate genes with a key role in the anthocyanin accumulation process in cranberry fruits.

P0138: Methods: High-throughput Methods

Same Day, Low Marker Density, High Throughput Genotyping

Heather Koshinsky, Maria Shin, Vineet Joshi and John D. Curry, Thermo Fisher Scientific, Santa Clara, CA

Genetic gain effectively relates four core factors that influence breeding progress: the degree of phenotypic variation present in a population, the probability that a trait will be transmitted from parent to offspring, the proportion of the population selected as parents for the next generation and the length of time necessary to complete a cycle selection. The length of time is not only how many generations are required to complete a selection cycle, but also how quickly the generations can be completed. This includes the time taken to obtain genotypes from samples.
Targeted genotyping by sequencing is emerging as a valuable tool for high throughput, low cost single nucleotide polymorphism (SNP) detection in both plant and animal genomics. Plant breeding has a long history of integrating the latest innovations in biology and genetics to enhance crop improvement. Using Eureka Genotyping Panels developed for various plants we have demonstrated genotypes in a day. Aquaculture breeding programs need rapid, high throughput genotyping to develop genetically improved stocks for cost-effective production. Older fish (more growing days), are often the larger fish. Thus, in aquaculture breeding the age of the fish can obscure genetic potential. Genotyping fish the same day they are spawned would increase the efficiency of the breeding program. Using a 500-plex Eureka Genotyping Panel developed for *Oncorhynchus keta* (chum salmon, a residual tetraploid), we have demonstrated genotypes in a day.

P0139: Methods: High-throughput Methods
Building High Quality, Chromosome-Scale, *de novo* Genome Assemblies by Scaffolding Next-Generation Sequencing Assemblies with Bionano Genome Mapping
Jian Wang, Andy Wing Chun Pang, Ernest Lam, Warren Andrews, Thomas Anantharaman, Alex R. Hastie, Henry Sadowski, Michael G. Saghbini, Zhanyang Zhu, Mike Austin, Han Cao and Mark Borodkin, Bionano Genomics, San Diego, CA
With the exception of a few model organisms, many biologically and economically important plants and animals still lack a reference-quality genome assembly that is crucial to the understanding of their biology. Their genomes are often complex and highly repetitive, making generation of high-quality assemblies almost impossible with next-generation sequencing (NGS) alone and without access to long-range structural information. Bionano's next-generation genome mapping (NGM) technology provides a solution to reconstruct the full genomic architecture of large and complex genomes.
Here, we present a novel direct enzymatic labeling approach which maintains the integrity of the DNA and allows us to create very contiguous Bionano maps which can then be used to scaffold NGS sequence assemblies to produce highly contiguous and structurally accurate hybrid assemblies that can span most repeat regions. This direct labeling method is compatible with a vast array of organisms
We validated our approach with the human NA12878 genome. Starting with NGS assemblies with N50 ranging from 0.18 – 0.9 Mbp, we produced hybrid assemblies with N50 from 70 to 80 Mbp. Chromosome-arm length scaffolds were assembled in 20 out of 23 chromosomes, and alignments show that they were consistent with the hg19 reference. The hybrid assemblies incorporated 80-90% of total NGS sequences with over 99% scaffolding accuracy. We will also show equally impressive scaffolds for a variety of plants and animals. For a low cost and only several days from sample-to-scaffold, this new method promises to set a new standard for genome finishing.

P0140: Methods: High-throughput Methods
Gentyane: A Service Platform for High Throughput Genotyping and Sequencing
Elodie Belmonte and Michaela Jody West, Gentyane INRA GDEC, Clermont Ferrand, France
Gentyane is a French platform offering services in the field of high throughput genomics. It is part of the GDEC (Genetics Diversity and Ecophysiology of Cereals) joint institute INRA-UCA in Clermont-Ferrand, France. The team is composed of 3 engineers, 1 engineer assistant and 3 technicians who are experts in genotyping and sequencing. The platform is certified ISO9001: 2008 and NF X 50-900 which ensures traceability and quality.
Our goal is to offer various panels of advanced technologies:

- Sequencing: PacBio Sequel (Data per SMRT Cell: 5Gb, reads up to 20kb)
- SNP genotyping: Affymetrix Axiom on GeneTitan (8 billon of data in 2016); Kaspar on Fluidigm Biomark (5 chips/day: 45600 data (96*96) and LightCycler480 (20 plates/day: 7680 data (plates 384))
- SSR genotyping: 3730XL (up to 16 plates 384 per day)
- High throughput DNA extraction on I7 Beckman (available in January)

The platform support research activities carried out in the different teams of the GDEC institute and performs genotyping and sequencing services to collaborators from either public or private sector. The experience gained over the years allows us to offer these services on a wide variety of species (animal or plant).

P0141: Methods: High-throughput Methods
Evolving Next Generation Sequencing for Production Agriculture: Increasing Throughput, Decreasing Effort and Delivering More
Jie Lu, Roy C. Willis, Angela Burrell, Michelle Swimley, Michelle Swimley, Christina Buchanan-Wright, Rick Conrad and Haktan Suren, Thermo Fisher Scientific, Austin, TX
The utility of restriction-enzyme genotyping by sequencing (GBS) in production agriculture is challenged because of the technology's limitations in SNP targeting and high rates of allele drop-out between samples. In contrast, targeted Genotyping by Sequencing (GBS) can deliver consistent, high marker call rates for specified SNPs in a high throughput, cost effective manner. AgriSeq™ targeted GBS allows up to 5000 markers to be simultaneously screened across 100s of samples in a single Ion Torrent sequencing run. The robustness of this technology has been demonstrated across 19 agriculturally relevant species, with marker call rates between 88-98%, >99% reproducibility and >99% concordance with orthogonal genotyping technologies. Here we report the expansion of the AgriSeq workflow to accommodate complex structural variants with in-dels ranging between 2-400,000bp.
Enhancements have also been made to the AgriSeq workflow to improve throughput. An additional 384 IonCode barcodes were developed to enable 768 sample multiplexing. Barcodes were validated over 3 different panels generating equivalent performance for ligation efficiency, uniformity and mapped reads. In addition, AgriSeq automation workflows were developed on the Gilson PipetMax to enable 384 samples processing with <1 hour of hands on time. Equivalent call rates, percentage on-target and coverage uniformity were observed between manual

and automated processing across two different panels, over 6 runs with 192 samples. The expansion of available barcodes and automation of the AgriSeq library prep enables up to 1536 samples to be processed each day while reducing operator fatigue, the potential for technical errors, and the sequencing cost per sample.

P0142: Methods: High-throughput Methods

Evaluation of an Actively Cooled, Cryogenic Homogenization Process on the Quality of HMW DNA

Raana Sheikh and David Burden, OPS Diagnostics, Lebanon, NJ

The need to isolate high molecular weight (HMW) DNA using high throughput (HT) methods has significantly increased with the rise of NextGen sequencing. Current HT methods include cryogenically bead beating tissues in vials/tubes held in solid aluminum pre-chilled blocks. During processing, however, the heat generated by grinding gradually increases the temperature of the samples, increasing the risk of DNA shearing.

Here, we review the application of a light-weight, actively-cooled holder (AC Block™) for cryogenic HT grinding. The AC Block™ contains an absorbent insert which, when charged with liquid nitrogen (LN2), emits LN2 vapors that actively cool samples during grinding. Also, compared to solid cryogenic blocks, the AC Block™ has less mass, resulting in less homogenizer motor wear and faster grinding ball acceleration and force.

A comparison of HMW DNA isolated from plant (corn) and animal (mouse) liver tissues cryogenically homogenized using the AC Block™, solid cryogenic blocks and mortar and pestle was performed. Tapestation results for animal tissue showed that greater than 48% of the DNA isolated using the AC Block™ was larger than 40 kb, compared to 27% using the solid aluminum block. Use of mortar and pestle resulted in 53% DNA over 40 kb, although yield was the lowest of all three methods. Plant samples processed using the same methods yielded less HMW DNA, though in roughly the same proportions. Overall, use of the AC Block™ resulted in less DNA shearing and greater yields of HMW DNA, compared to alternative methods.

P0143: Methods: High-throughput Methods

Is the High-Throughput Human OpenArray® System Useful for Profiling MicroRNAs in Melanoma Regression in a Swine Model?

Claudia Bevilacqua[1], Fany Blanc[2], Marie-Noelle Rossignol[3], Guillaume Piton[4], Sylvain Marthey[5] and Silvia Vincent-Naulleau[4], (1)GABI, Plateforme @BRIDGE, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France, (2)GABI, Genetic-Immunity-Health team, INRA, AgroParisTech, Université Paris-Saclay, Jouy en Josas, France, (3)INRA, JOUY EN JOSAS, France, (4)CEA, Fundamental Research Direction (DRF), Cellular and Molecular Radiobiology Institut (IRCM), Université Paris-Saclay, LREG, Jouy en Josas, France, (5)UMR GABI, INRA, Université Paris Saclay, Jouy en Josas, France

Cutaneous Melanoma (CM) is the most aggressive cancer of the skin due to its high metastatic potential. Despite recent encouraging results from target therapies, treatment of invasive melanoma is still challenging and new data on melanoma biology are required. MicroRNA (miRNA) are post-transcriptional regulators of gene expression. Altered miRNA expression profiles and the identification of their targets in cancer imply the high potential of miRNA as diagnostic and prognostic markers for development of new therapeutic strategies. The Melanoma-bearing Libechov Minipig (MeLiM) model is a relevant animal model of CM since melanocytic lesions are similar to human counterparts. MeLiM pigs are born with CM that progress to spontaneous regression leading to a high rate of survival even in metastasis-bearing animals. In this study, we investigated whether miRNA could be involved in this fascinating process.

Even if the release of the pig genome sequence in 2012 was a major step for using pig models for biomedical research, some data are always limited in swine. For example, in miRbase 21, only 382 sequences from Sus scrofa miRNA precursors are listed against 1881 for Homo sapiens. For this screening, we tested the use of TaqMan® OpenArray® human MicroRNA Panels (Pool A and B, ThermoFisher) charged with 750 human miRNA to compare quickly and easily, miRNA profiles between progressive and regressive MeLiM tumors. With 60% of positive amplification, results show that this approach could be an alternative of miRNA sequencing for miRNA screening in some species where "large" dedicated tools are missing.

P0144: Methods: High-throughput Methods

A Novel High Throughput Microbiome System for Cultivating and Screening Bacteria

Alexander James Hallock, GALT, Inc., San Francisco, CA

Advances in microbiome research over the past decade have yielded a cascade of information that has significantly increased the understanding of the role that bacteria and bacterial communities play in human health, agriculture, and other environmental sciences. However, a deeper understanding of the role that microbes play has been hampered by outmoded laboratory tools currently used to cultivate and screen bacteria prior to more advanced analysis of their function. Here we describe a system that is capable of massively parallel cultivation, and screening of large numbers of bacteria from any sample source. The application of a high density micro fabricated array, and a highly accurate and precise desktop instrument has been demonstrated to dramatically improve throughput and workflow productivity. The technology has been successfully applied to cultivating and growing bacteria from a variety of sample sources, as well as high throughput direct screening of microbiome samples for specific phenotypic function.

P0145: Methods: High-throughput Methods

Nugen's Allegro™ Targeted Genotyping: An Accurate and Cost-Effective Sequencing Workflow for any Genome.

Michael T Lovci, Stephanie C Bruns, Marie Eide, Luke Sherlin and Joe Don Heath, NuGEN Technologies, San Carlos, CA

Allegro™ Targeted Genotyping allows interrogation of DNA heterogeneity at an industrial scale. By using NuGEN's patented Single Primer Enrichment Technology (SPET) we provide strand-specific readouts of genome sequence, enabling sensitive detection of allele frequency for as many as hundreds of thousands of sites. Our custom panel design process is a complimentary concierge service that will provide designs that meet the specific business or scientific needs of your study. Raw data are amenable to industry-standard genotyping-by-sequencing analysis pipelines. Additionally, we provide the ability to perform genotyping through a cloud analysis platform built for Allegro™ customers. Here, we

present a validation study using an oligo panel targeting human SNPs. Reference gDNA provided by Horizon Diagnostics with a known allele mixture is used as a gold-standard; we observe high concordance with known minor-allele frequencies as low as 1% (pearson R > .98). Our process grants massive cost-reductions compared to whole-genome assays at a price-point comparable to or better than microarray approaches with the fidelity and scalability benefits of Next Generation Sequencing.

P0146: Methods: High-throughput Methods
Fast Ordered Sampling of DNA Sequence Variants
Anthony Greenberg, Bayesic Research, Ithaca, NY
While the amount of genomic data available is growing quickly, it is matched by increasing power and storage space of consumer-grade computers. Even applications and pipelines that require powerful servers can be quickly tested on desktop or laptop computers if we can generate representative samples from large data sets. Such subsets are also useful for statistical applications, such as repeated re-sampling. A fast and memory-efficient method that preserves the order of the original records would thus find many applications. A sampling method developed for tape drives 30 years ago appears to fit the requirements. I implement a version of this technique for files containing genetic variant genotype data. I test its performance on modern solid-state and spinning hard-drives, and show that it performs well compared to a simple sampling scheme. I illustrate the utility of the technique by developing a sampling-based method to quickly estimate genome-wide patterns of linkage disequilibrium (LD) decay with between-variant distance. I provide open-source stand-alone software that samples loci from several variant format files, a separate program that performs LD decay estimates, and a C++ library that lets developers incorporate these methods into their own projects.

P0147: Methods: High-throughput Methods
Classification of Wheat Varieties in Satellite Images to Perform GWAS
Samuel R. Revolinski and Zhiwu Zhang, Dept. of Crop and Soil Science, Washington State University, Pullman, WA
With dramatic reduction of sequencing cost, field trials have become the major burden of plant breeding to improve grain yield, quality and resistance to biotic and abiotic stress, especially on large scale. Abundant available satellite images have the potential to provide valuable data in regards to infection from disease, responses to drought and heat, as well as grain yield and quality. Most of these are due to the improvement of sensing resolution and shorter time intervals between pictures taken. The Satellite WorlView-2, launched in 2009, provides sensing at resolution within a meter on a near daily base. We review the satellite imagery resources and usages, including disease detection, crop discrimination, drought tolerance, nitrogen efficiency. Our objective was to inspire the joint usage of remote imagery data and genomic data for the genetic improvement in crops.

P0148: Methods: High-throughput Methods
PubRunner: A Tool for NLP Researchers, now used to Unite the PubMed and Agricola Abstract Corpora
Jake Lever, University of British Columbia, Vancouver, BC, Canada and **Ben Busby**, NCBI, Bethesda, MD
Modern biology research relies on structured data contained within biological databases. These databases contain information about genetic interactions, pathogens and other associations. Text mining tools offer an opportunity to extract knowledge in order to build these databases without the incredible time and cost associated with manual curation. Unfortunately most efforts have focussed on abstracts from PubMed. Many journals that are of value to plant biology researchers are not commonly used in text mining. To this end, we look to AGRICOLA, a long-term resource managed by the US Department of Agriculture, that contains a large number of abstracts from plant biology and agriculture journals. We evaluate the integration of AGRICOLA into the PubRunner project which is a framework to allow easy application of text mining tools to large corpora. By adding in MarcXML converters, we easily add AGRICOLA to an existing text mining tool that calculate word vector representations. To further show the use of AGRICOLA in text mining applications, we develop a nutrigenomics analysis using AGRICOLA and PubMed. This identifies cooccurrences of genes with food items in abstracts and is a valuable tool to explore the genetic interactions of food. We hope this analysis will show the value of text mining beyond PubMed for plant biology information.

P0149: Methods: High-throughput Methods
PolySNP: An R Package for Calling Polyploid SNP Array Data using Gaussian Mixture Models
Matthew T. McGowan[1], Zhiwu Zhang[2] and Stephen P. Ficklin[1], (1)Washington State University, Pullman, WA, (2)Dept. of Crop and Soil Science, Washington State University, Pullman, WA
The genomic complexity in polyploid plant species makes genotype calling difficult when processing data generated from a high-density SNP array. We present a novel computational method and a software package for calling genotypes from raw hexaploid wheat data generated using the 9k iSelect® assay previously developed for wheat. This method involves fitting an indeterminate number of Gaussian mixture components and identifying the optimal number of clusters using an EM-like algorithm implemented in the 'Rmixmod' package. Then markers with bi-allelic patterns are further analyzed by merging outlier clusters and identifying heterozygous clusters. Genotypes are then called based on cluster assignment. Furthermore, models generated with a diverse population can be later used to call genotypes for smaller populations, drastically reducing computational complexity for subsequent calls. This method was tested using a diverse wheat population (n = 1654) and resulting genotypes were compared to previously called genotypes using the current standard method of manual curation. Genomic predictions were generated for both genotype sets using the gBLUP method implemented in the 'rrBLUP' package in R for five different phenotypes. Regression coefficients for predicted vs observed values were improved by 1.38% when using genotypes generated with this new method. Despite an increased computational cost of using Gaussian mixture models, a reduced supervision requirement and increased ability to resolve complex signal patterns allow it to generate more predictive genotypes with less manual manipulation.

P0150: Methods: High-throughput Methods
Riptide™ Ultra-High Throughput – Rapid Library Preparation for NGS

Keith Brown[1], Azeem Siddique[2], Steve Head[2], Phil Ordhoukhanian[2], Jorge Bajena[2], Nils Homer[3] and Andrew Su[2], (1)iGenomX, Carlsbad, CA, (2)The Scripps Research Institute, La Jolla, CA, (3)Fulcrum Genomics, Somerville, MA
Whole Genome Shotgun Sequencing has become the tool of choice for microbial genome analysis. Rapidly declining costs of sequencing, data analysis, data storage and database access will continue to drive adoption. Library construction has not kept pace with these advancements, with costs of preparing a next generation sequencing (NGS) library often exceeding the cost of sequencing. Popular methods of library construction for NGS include fragmentation, end-repair and adapter ligation, or transposase-mediated adapter insertion. The Riptideä High Throughput DNA Library Prep is distinctly different in its approach because it relies on polymerase-mediated primer extension and termination for library preparation. The initial step of the prep, involving primer extension with barcoded random primers, is performed in a 96-well plate. Each well of the plate contains primers with a unique barcode; consequently, the library generated from each well is uniquely identifiable and can be bioinformatically traced back to the original sample after sequencing. Following this step, the primer extension products are combined into one pool and all subsequent steps, including second strand synthesis and PCR, are performed with the single pool. The library prep is fast, easily automatable and can be tuned to genomes of high and low GC content. With automation, 960 samples can be easily processed in a single day. The technology will aid genetic research by helping to increase sample throughput and by reducing processing steps and operating costs. Data is presented here showing its application on a diverse set of microbial genomes.

P0151: Methods: High-throughput Methods

A Robust and Versatile Statistical Genetic Pipeline for Plant Breeding

Fred A. van Eeuwijk[1], Martin P. Boer[1], Chaozhi Zheng[1], Marcos Malosetti[1], Willem Kruijer[1], Bart-Jan van Rossum[1], Maikel Verouden[1], Darren A. Murray[2], Allen R Green[2], David Marshall[3] and Iain Milne[3], (1)Wageningen University & Research - Biometris, Wageningen, Netherlands, (2)VSN International Ltd, Hemel Hempstead, United Kingdom, (3)The James Hutton Institute, Dundee, United Kingdom
Within plant breeding new data generating techniques in the field of genotyping, phenotyping, omics techniques and envirotyping present new opportunities for increasing the efficiency of breeding programmes. To take advantage of these developments new statistical techniques and software are required. We present a pipeline containing modules for 1) data quality control at all stages of statistical analysis relevant to breeding programmes, 2) phenotypic analysis of single and multiple trials, 3) genotype by environment interaction, 4) population genetic analyses, 5) imputation, 6) genetic map construction for bi- and multi-parental populations and LD analysis, 7) QTL mapping for biparental and multiparental populations, 8) QTL mapping/GWAS/genomic prediction for single and multiple traits and environments, 9) decision support for selection of superior genotypes and crossing partners, 10) visualization tools to support reporting and decision making. New algorithms and methods are included for spatial analysis of field trials by two-dimensional splines, the generation of conditional QTL genotype probabilities for all kinds of breeding populations by a continuous time Markov process, used in genetic map construction and QTL mapping, and mixed model estimation algorithms for multi-trait analysis. The pipeline is based on R procedures that call on asreml-R mixed model procedures when necessary. The purpose of the pipeline is to make state of the art statistical and decision support methods accessible to a wide audience of breeders and geneticists via an intuitive user interface. The modularity of the pipeline allows easy adaptation to specific user requirements.

P0152: Methods: Markers

Multiplex Restriction Amplicon Sequencing (MRASeq), a New Next Generation Sequencing-Based Marker Platform for Genotyping

Amy N. Bernardo, Kansas State University, Manhattan, KS, Paul St Amand, USDA-ARS, Hard Winter Wheat Genetics Research Unit, Manhattan, KS and **Guihua Bai**, Hard Winter Wheat Genetics Research Unit, USDA-ARS, Manhattan, KS
Marker-assisted breeding enables the indirect selection of traits that are difficult and/or costly to phenotype thereby saving time and money, and increasing selection efficiency. To be useful in breeding programs, markers for genome-wide genotyping must be low cost, randomly distributed throughout the genome, high-throughput, and technically simple. We developed a PCR and NGS-based, low cost, high-throughput genotyping technology for genome-wide marker assays. This technology, designated as Multiplex Restriction Amplicon Sequencing (MRASeq), reduces genome complexity by PCR-amplification of selected portions of genomic regions flanked by restriction sites and is achieved using tailed and semi-degenerate PCR primers with restriction enzyme sequence at the 3'-end. MRASeq is flexible because the restriction enzyme sequence and the adjacent degenerate base sequence in the primers can be altered to suit the species of interest. MRASeq uses restriction sites as primer sites and does not make use of restriction enzymes. The incorporation of unique barcodes during a second PCR allows hundreds of samples to be multiplexed in one sequencing run. Linkage mapping of polymorphic MRASeq SNP markers in an allohexaploid wheat biparental population showed random distribution of SNPs across genomes. MRASeq on wheat and barley natural populations generated thousands of SNPs suitable for genomic selection. Therefore, this marker platform can be used for linkage mapping, background selection, or any other purpose in which large numbers of markers are needed. This simple, flexible and high-throughput genotyping method should be useful in genotyping laboratories, plant breeding programs, and genetic research.

P0153: Methods: Markers

New Genotyping Technology, GRAS-Di, Using Next Generation Sequencer

Hiroyuki Enoki and Yoshie Takeuchi, Toyota Motor Corporation, Aichi, Japan
We developed new genotyping technology, Genotyping by Random Amplicon Sequencing-Direct (GRAS-Di). This technology consisted of sample preparation using high concentration random primer, NGS and data analysis. The sample preparation was very simple. It was not necessary to do primer design, enzyme digestion, fragmentation, size selection, adaptor ligation, and sample normalization. It was only two steps PCR for NGS library without specialized equipment. Rice BIL population was used for evaluation of genotyping by GRAS-Di (96 samples / lane of HiSeq2500). The number of reads for each amplicon was highly reproducibility, $r > 0.99$, with repetition. Over ten thousand SNPs were detected among the BIL population and the SNPs were distributed uniformly rice genome. The ratio of missing value was very low, 1.5%. The reproducibility of SNP was 99.9% with repetition. If there was no reference sequence, genotype data could be detected by GRAS-Di

using original algorism based on amplicon analysis. Theoretically, the technology is also applicable to other creatures, including highly polyploidy creatures. We performed the applicability test for several creatures. The result shown that the technology was applicable for over fifty creatures, including wheat, soybean, tomato, potato, sugarcane, cow, pig, chicken, tuna and human. The technology could be provided over 30,000 multiplex sequencing at once. We think that GRAS-Di would be very easy and very powerful technology for genome wide genotyping in many creatures. We signed licensing agreement with Kazusa DNA Research Institute, Eurofins Genomics, and GeneBay for GRAS-Di.

P0154: Methods: Markers

Genotyping-by-Sequencing: Applications from Conservation Genetics to Genomic Selection

Jeanne Jacobs[1], Shannon Clarke[2], Marty Faville[3], Andrew Griffiths[3], Mingshu Cao[3], Rachel Tan[3], Tracey Van Stijn[2], Rayna Anderson[2], Rachael Ashby[2], Suzanne Rowe[2], Timothy Bilton[2], Rudiger Brauning[2], Alan McCulloch[2], Stephen Goldson[1], Ken Dodds[2] and John McEwan[2], (1)AgResearch, Lincoln Science Centre, Christchurch, New Zealand, (2)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (3)AgResearch, Grasslands Research Centre, Palmerston North, New Zealand

Genotyping-by-Sequencing (GBS) is used for generating high-density genetic marker data in a cost-effective manner across a range of applications in biology. We have developed the infrastructure and skill base required to apply (high throughput) GBS for different purposes across a wide range of species. Thus far, we have optimised GBS in 50 different species encompassing plants, mammals, shellfish, fish, birds, and insects. Several of these species have no prior genetic and/or genomic information, making GBS the method of choice. The methodology has been scaled from small sample sizes (< 100) for diversity studies, up to many thousands in animal and plant breeding programmes where GBS underpins genomic selection. We have used GBS for environmental metagenomics, population genetics and conservation genetics studies. Continuous improvements in wet-lab methods have enabled increased quality and quantity of data generated. Data analysis has been enhanced through improved bioinformatic workflows tailored to each species, including statistical methods designed specifically for (low depth) GBS data such as 'KGD' for developing genomic relationship matrices (Dodds et al., BMC Genomics 16:1047, 2015) and 'GUSMap' for constructing linkage maps. Components of the data analysis workflow are available in a public Github repository (https://github.com/Agresearch). Many studies are in collaboration with universities & research organisations, as well as commercial entities, both nationally and internationally.

P0155: Methods: Markers

CandiSNP and CHERIPIC: Tools for Rapid Fine Mapping of Causative Mutations using Draft or Completed Genome Sequences.

Ghanasyam Rallapalli, University of East Anglia, Norwich, United Kingdom

Traditional Map Based Cloning approaches, used for the identification of desirable alleles, are extremely labor intensive and years can elapse between the mutagenesis and the detection of the polymorphism. Combining bulk segregant analysis with high throughput sequencing (HTS), referred to as Mapping-By-Sequencing (MBS) has accelerated the identification of causative mutations for genomes with good quality reference sequences. Our group's CandiSNP analysis and visualisation application, generates density plots and table with a list of candidate causative mutations, defined as SNPs causing non-synonymous changes in annotated coding regions using SNPs obtained from bulk segregant sequencing (BSS) data. CandiSNP is a user-friendly application that will aid in novel discoveries from forward-genetic mutant screens. The web-application is freely available online at http://candisnp.tsl.ac.uk/ and source code at http://github.com/TeamMacLean/candisnp/. Although MBS studies are effective, require an ordered genome assembly and cannot be used with fragmented, un-scaffolded draft genomes, limiting their application to model species and precluding many important organisms. We developed a computational method and software implementations to address this gap in resource. Called CHERIPIC (Computing Homozygosity Enriched Regions In genomes to Prioritise Identification of Candidate variants - http://cheripic.tsl.ac.uk/ and https://github.com/shyamrallapalli/cheripic/), it makes use of fragmented genome assemblies resulting from BSS data to call variants and identifies a causative mutation or a few closely linked variants that help narrow down the region harbouring the trait of interest. CHERIPIC has been applied to assemblies of bulked whole genome sequence data from Arabidopsis, bulked RNA-seq data from maize, bulked exome data of barley and identified variants that are very closely linked to the region of the causative mutation.

P0156: Methods: Markers

Development of Sets of SNP Markers for Germplasm Characterisation and Cultivar Identification in Fruit Trees and Berries

Werner Howad, Iban Eduardo, Mourad Mnejja, Jose De la Cruz, Kostas Alexiou, Amparo Monfort and Pere Arus, IRTA, Center for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Bellaterra (Barcelona), Spain

Developing robust and reliable sets of molecular markers for germplasm characterisation and cultivar identification is a challenging task especially in polyploid species. We set up a pipeline including different approaches using peach and apricot as model species. In peach public available SNP data (IPSC 9K SNP chip) were used to select the most suitable SNPs based on MAF values and map position, whereas in apricot publicly available SNP and sequence data were complemented with in-house produced sequence data. For each of the two species the TaqMan® OpenArray® technology was chosen to design 64 SNP assays, which were evaluated in hundreds of varieties. Final sets of 55 SNPs (peach) and 54 SNPs (apricot) have been proven to be suitable and are in routine use now. For diploid and octoploid strawberry the peach approach was applied using data of the 90 K Axiom® SNP array to select a set of 64 SNPs. After evaluation of the OpenArray® assay a final set of 55 SNPs is now in use. Due to the limited data available in blueberry and raspberry we decided to design the 64 SNP set assays based on in-house produced sequence data obtained from 8 varieties per species under the criteria established in the apricot approach. These SNP assays will be evaluated with an elevated number of varieties to select the final sets for routine use. The SNP pipeline has also been successfully applied for the development of sets of SNPs in diploid and polyploid cereal species.

P0157: Methods: Markers

Towards the Development of a Selection Panel for Berry Size in Table Grapes Based on the Use of SNP and Indel Markers
Claudia Muñoz-Espinoza[1], José Correa[1], Matías Burgos[1], Thierry Lacombe[2], Valerie Laucou[2], Jean-Pierre Péros[2] and Patricio Hinrichsen[1], (1)INIA, Centro de Investigación La Platina, Santiago, Chile, (2)AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
Berry size is considered as one of the main selection criteria in table grapes breeding programs, due to consumer´s preferences. Considering its economical importance, it is relevant to determine the genetic architecture of this trait and elucidate the mechanisms involved in its expression. To approach this issue, 30 SNPs and eight INDELs were identified using a transcriptomic approach (RNA-Seq), found among segregants from a 'Ruby' x 'Sultanina' crossing with contrasting phenotypes for berry size. The ultimate aim of this work is to develop a selection tool, which can be included as a regular protocol in breeding programs. In order to select the best combination of candidate polymorphisms as predictor markers, two groups of varieties with different genetic backgrounds were considered. The first consisted in 30 table grape varieties representing the cultivar diversity in Chile. This group was phenotyped for berry size, seed content as well as polar and equatorial diameter during three seasons at La Platina Research Station. The second group included 113 varieties belonging to the INRA-Vassal collection for *V. vinifera* (France), selected for their contrasting berry sizes. Preliminary genotype-phenotype association results using Tassel software, unbalanced ANOVA and *Random Forest* analyses including segregants and table grapes varieties showed several degrees of association of these markers with berry size (10.6% to 34.5%). Both collections have been genotyped using candidate markers with higher association to berry weight and a High Resolution Melting (qPCR-HRM) platform, to assess the usefulness of these polymorphisms as selection markers.
Financed by FONDECYT-Chile grants 3150519 and 1171378.

P0158: Methods: Markers
Gene Discovery for Strawberry Flavor
Chris Barbey[1], Kevin M. Folta[1], Vance M. Whitaker[2] and Sujeet Verma[2], (1)University of Florida, Gainesville, FL, (2)University of Florida, Wimauma, FL
Associating crop traits with their causative genetics is crucial for agricultural improvement. This deductive process is challenging in cultivated strawberry due to an unusually complex auto-allooctoploid genome. Contemporary sequencing techniques typically have insufficient resolution to discriminate octoploid strawberry's numerous allelic variants. Two strawberry traits of broad importance are improved flavor and disease resistance. To identify candidate genes controlling these traits, high-resolution genomics assays were applied to detect sequence variations at the subgenomic level. For associating flavor phenotypes, fruit volatile metabolomes were rigorously derived from 263 individuals using a series of statistical alignment techniques on non-targeted GC/MS data. Over 25,000 subgenomic sequence variants were tracked through multiple generations using a specially designed octoploid strawberry SNP-array. These inherited sequence variants were successfully correlated with the production of various strawberry flavor and aroma volatile compounds. Among these, Linalool (floral-aroma), Methyl Anthranilate (grape-aroma), and Methyl 2-Hexenoate (pineapple-aroma) and others were genetically associated with high confidence and precision. Demonstrating the validity of these results, linalool was mapped precisely to the locus of its known biosynthesis allele. In a similar approach, high-variance transcripts from 65 fruit transcriptomes were successfully correlated to their genotypes, indicating that differential expression of over ninety-five fruit transcripts is due to segregating genetic *cis* and *trans* factors (eQTL) .

P0159: Methods: Markers
Development of Genomic Simple Sequence Repeat Markers for *Atkinsonella hypoxylon* (Peck) Diehl.
Chandra Sekhar Thammina[1], Keenan Amundsen[2], Matthew H. Kramer[3] and Scott E. Warnke[1], (1)USDA-ARS-US National Arboretum, Beltsville, MD, (2)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, (3)USDA-ARS, Beltsville, MD
Atkinsonella hypoxylon (Peck) Diehl. is an Ascomycete fungus belonging to the tribe Balansieae (Clavicipitaceae). It grows as an epiphyte on the grass *Danthonia spicata* (L.) Beauv. *D. spicata* is often found in low input turfgrass areas on the East Coast of the United States and has the potential for development as a new native low input turfgrass species. Published research suggests that endophytic and epiphytic fungi can improve the tolerance of host plants to biotic and abiotic stresses. The objective of this study was to develop polymorphic genomic-SSR primer pairs and utilize them to assess the levels of variability between *A. hypoxylon* isolates collected from different locations. Ninety-six trinucleotide genomic-SSR primer pairs were tested on *A. hypoxylon* isolates using high resolution melt analysis. Later, 30 primer pairs were selected based on their amplification profiles and the PCR products were sequenced resulting in 11 primer pairs amplifying polymorphic loci compared to the source contig sequence. In the future, we will utilize these polymorphic markers to analyze the genetic diversity and study phylogenetic relationships of *A. hypoxylon* isolates collected from *D. spicata* plants growing in different areas of US with a final goal of establishing the role of *A. hypoxylon* in the evolution of reproductive mechanisms in *D. spicata*.

P0160: Methods: Markers
Development and Characterization of Microsatellite Markers on Linkage Group 6 of European Hazelnut
Golnaz Komaei Koma and Shawn A. Mehlenbacher, Oregon State University, Corvallis, OR
Advances in next-generation sequencing technology have accelerated the development of microsatellite markers, also known as simple sequence repeats (SSRs). In this study, the genome sequence of 'Jefferson' hazelnut (*Corylus avellana* L.) assembled from Pacific Biosciences (PacBio) and Illumina sequences was used as the reference genome. Genomic DNA of seven other cultivars was sequenced using Illumina. Tablet software was used for *in-silico* comparisons of the reads of the seven cultivars aligned with the 'Jefferson' reference sequence and identification of polymorphic SSRs. Eastern filbert blight is a serious threat to the hazelnut industry in Oregon. Resistance from 'Gasaway' is conferred by a dominant allele at a single locus on linkage group 6. Bacterial artificial chromosome (BAC), BAC end and molecular marker sequences were used to identify PacBio contigs in this region. Four contigs were selected and di-, tri- and tetra-nucleotide repeats were identified. Removal of repeats containing only A's and T's reduced the number of unique fragments to 451. Tablet software was used to visualize the aligned reads, and identify 116 polymorphic SSRs for which primers were designed. PCR amplification and electrophoresis on agarose gels confirmed that 60 were polymorphic. The parents of the mapping population and 48 diverse accessions were genotyped and allele

sizes determined using GeneMapper software. The data were analyzed using PowerMarker and Cervus software. The number of alleles per locus ranged from 5 to 27 with the average of 13.62. The mean observed heterozygosity, expected heterozygosity, and polymorphism information content were 0.85, 0.82 and 0.79, respectively.

P0161: Methods: Markers

Target Gene Screening and Application to Watermelon Breeding

Geung-Joo Lee, Chungnam National University, Daejeon, South Korea, Saminathan Subburaj, Chungnam National University, Daejeon, Korea, Republic of (South) and Hyounsub Lim, Chungnam National Univeristy, Daejeon, South Korea

In horticultural crops, various colored cultivars accumulated different major or minor carotenoids. Carotenoid content in fruit is highly influenced by carotenoid biosynthesis pathway and its related genes. An investigation of transcriptional expression of key carotenoid-related genes may serve to identify and understand the regulation of fruit development in watermelon. One of the strategies to increase the carotenoid yield is to explore the target genes in the pathway and to find elicitors or factors (cultural or genetic) to maximize the expression of the genes. Large number of transcription factors, particularly from tomato, have been shown to affect carotenoid accumulation through the regulation of fruit ripening. These transcription factors include *RIN*, *TAGL1*, *AP2a*, *ERF6*, *DET1*, *APRR2*-Like, *SGR*, *BZR1-1D*, *PIF-1* etc. A total of 1,444 TFs are available in watermelon genome. Finding TF- motif binding sites in promoters of carotenoid biosynthetic genes would helps to explain the transcriptional Regulation of fruit ripening process in watermelon. Also determination of allele specific variations in 5' UTR regions in the carotenoid biosynthesis rendered us to develop reliable DNA marker to discriminating LCYB alleles for red and yellow flesh fruits. Large number of transcriptional factors including RIN, SISGR1, PIF-1, AP2a, ERF6, etc was found to be positively or negatively interacting with some structural genes in the carotenoid pathway. Recently CRISPR/Cas9 system have been applied to mutate the function of *phytoene desaturase* (PDS) genes and *RIN* TF. Therefore, It is possible to use CRISPR/Cas9 system in watermelon cultivars to improve the levels of Lycopene in watermelon fruit by mutate the key negative regulatory genes or TFs which controls the genes involves in carotenoid biosynthesis pathway genes.

P0162: Methods: Markers

Peanut Germplasm Screening for Aflatoxin Accumulation and Genetic Fingerprinting

Renee S. Arias[1], Victor S. Sobolev[1], Guillermo J. Seijo[2], Brian E. Scheffler[3] and Alicia N Massa[4], (1)USDA-ARS National Peanut Research Laboratory, Dawson, GA, (2)Instituto de Botanica del Nordeste, Corrientes, Argentina, (3)USDA-ARS, Stoneville, MS, (4)USDA-ARS-National Peanut Research Laboratory, Dawson, GA

Aflatoxin contamination in peanut seeds is still a serious problem for the industry and human health. There are no aflatoxin resistant cultivars, and given the narrow genetic background of cultivated peanuts, search for resistance is now focused on wild species. We adapted our aflatoxin accumulation testing method to the small-size seed of 20 wild peanut species, used only four seeds per accession, tested viability of each seed tested, and quantified the main four aflatoxins A_1, A_2, B_1 and B_2 for each seed using ultra-performance liquid chromatography (UPLC). In parallel, we fingerprinted each accession using 373 simple-sequence repeat (SSR) (288 novel) and Insertion-Deletion (InDel) markers to keep a genetic record of the accessions tested and proper identification within the germplasm collection. Levels of aflatoxin observed among the peanut species tested varied from 0 - 14000 $ng.g^{-1}$ and 155 $ng.g^{-1}$ of aflatoxin B_1 and B_2, respectively. Multivariate analysis by Neighbor-Joining, Principal Component Analysis and Principal Coordinate Analysis using 134 (36 %) transferable markers, in general grouped the samples according to their reported genomes. The best 88 markers, based on high fluorescence, good scorability and transferability, are reported. UPIC program identified a group of three high quality markers as sufficient to discriminate all 20 species tested. High quality markers with high UPIC scores and their corresponding significant hits on BLAST2GO are also reported. These tools can be used for the systematic search of aflatoxin resistant germplasm keeping record of the genetic fingerprinting of the accessions tested for breeding purpose.

P0163: Methods: Markers

Development of Coronatine-Insensitive 1 (COI 1) Markers for Mining Glandular Trichome-Based Resistance to Coconut Scale Insect

Roanne R. Gardoce[1], Darlon V. Lantican[1], Melvin P. Dancel[1], Frenzee Kroeizha L. Pammit[1], Vangelene T. Linga[1], Susan R. Strickler[2], Lukas A. Mueller[2] and Hayde F. Galvez[1], (1)University of the Philippines Los Baños, College, Laguna, Philippines, (2)Boyce Thompson Institute, Ithaca, NY

Trichomes form the first line-of-defense against herbivorous insects and pathogens. Glandular trichomes are good subjects for resistance studies as they produce a wide variety of chemical compounds that deter, immobilize or toxify insect pests and pathogens. A study suggested the role of coronatine-insensitive 1 (COI 1) in jasmonic acid signaling processes for glandular trichome-based insect resistance in tomato which possesses type-IV glandular trichomes like the coconut.

The aim of this study is to develop markers that can be used to detect natural variants in in COI 1 to find a natural source of resistance to the coconut scale insect (*Aspidiotus* sp.). COI 1 gene sequences of *Elaeis guineensis* L., a close relative of coconut, were retrieved from NCBI, and used to find homologs in a draft coconut genome assembly. Three contigs (3054 bp, 2331 bp, and 1833 bp) were found in the 'Catigan Dwarf' variety coconut draft genome sequence. Twelve (12) primers had PCR amplification in coconut genomic DNA. Homology analysis revealed that the regions amplified by four primers share homology to known COI 1 and COI 1-like sequence. A set consisting of candidate resistant coconuts, controls and 73 varieties/accessions from the Philippine Coconut Authority field genebank was constituted for mining natural gene sequence variations through Eco-TILLING approach. Results showed that natural SNPs were identified from five (5) COI 1 markers. The full length mRNA of COI 1 in coconut will be cloned and characterized. Linked SSR markers are also currently being designed.

P0164: Methods: Markers

Comparison of Three PCR-Based Methods for SNP Genotyping in Sugar Beet

Chiara Broccanello[1], Andy Funk[2], Claudia Chiodi[1] and **Piergiorgio Stevanato**[1], (1)University of Padova, Legnaro (Padova), Italy, (2)Michigan State University, East Lansing, MI

PCR allelic discrimination technologies have broad applications in the detection of single nucleotide polymorphisms (SNPs) in genetics and genomics. The use of fluorescence-tagged probes is the leading method for targeted SNP detection, but assay costs and error rates could be improved to increase genotyping efficiency. A new assay using RNAse-H PCR (rhPCR) attempts to reduce error rates from primer dimers while lowering costs compared to existing technologies. Before rhPCR can be widely adopted, it is important to validate its effectiveness versus established methods. The aim of this study was to compare the accuracy, sensitivity and costs of three PCR-based, high-throughput SNP genotyping approaches; TaqMan, KASP, and rhPCR. For each approach, assays were designed to genotype 33 SNPs in a set of 96 sugar beet individuals obtained from 12 genotypes. The sensitivity of each assay was tested using a series of 20 dilutions from 0.1 ng to 100 ng per reaction. Reactions were carried out on the QuantStudio 12K Flex Real-Time PCR System (Life Technologies, CA, USA). The call-rate, defined as the percentage of genotype calls relative to the possible number of calls, was 97%, 97.6%, and 98.1% for TaqMan, KASP, and rhPCR, respectively. Discordance among SNP calls was restricted to 9 of the 33 SNPs, with discordance between methods ranging from 0.11% to 0.66%. The sensitivity test demonstrated that the limit of detection (LOD) of rhPCR was the lowest of the three assays (0.2 ng of DNA per reaction). Costs per reaction were Euro 0.29 for TaqMan, Euro 0.11 for KASP, and Euro 0.13 for rhPCR. In conclusion, rhPCR produced slightly more calls than either TaqMan or KASP while remaining competitive in terms of cost per SNP.

P0165: Methods: Other Genome Methodology

QTLseqr: An R Package for Bulk Segregant Analysis with Next-Generation Sequencing

Ben N. Mansfeld, Michigan State University, East Lansing, MI and Rebecca Grumet, Department of Horticulture, Michigan State University, East Lansing, MI

Since the early 1990's, Bulk Segregant Analysis (BSA) has been a valuable tool for rapidly identifying markers in a genomic region associated with a trait of interest. BSA is amenable to any type of codominant markers, including single nucleotide polymorphism (SNP) markers. This has allowed for the adaptation of this technology for use with next-generation sequencing (NGS) reads. SNPs detected in reads aligning to genomic regions closely linked to the trait should deviate from the expected ~50% representation observed in non-linked regions. In the past several years, the main pipeline used for NGS-BSA for plant breeding research was QTL-seq. While this approach has been widely used in several crops for many traits, the released pipeline has not been updated in several years, and as a result software and version incompatibility issues have arisen. This limits the widespread utilization of this otherwise well-designed pipeline. While an alternate approach for evaluating statistical significance of QTL from NGS-BSA based on a tricube-smoothed G statistic exists, a software implementation was never developed or distributed. We thus present "*QTLseqr*", an R package for NGS-BSA, that incorporates both methods above. QTLseqr, can quickly import and filter SNP data from the Genome Analysis Tool Kit (GATK) pipeline, then calculate and plot SNP distributions, relative allele frequencies, the tricube-smoothed G values, as well as $\log_{10}$(p-values). This allows for easy plotting and identification of QTL regions. QTLseqr is currently available at https://github.com/bmansfeld/QTLseqr

P0166: Methods: Other Genome Methodology

Absolute Quantification of Genetically Engineered Canola Events By Digital PCR Is Affected By the Type of Reference Gene Used

Tigst Demeke and Monika Eng, Canadian Grain Commission, Winnipeg, MB, Canada

Droplet digital PCR (ddPCR) has been used for absolute quantification of genetically engineered (GE) events. One of the main advantages of ddPCR over real-time quantitative PCR is that there is no need to use reference materials and standard curve. On the other hand, absolute quantification of GE events using duplex ddPCR requires the use of appropriate target and reference gene sequences (for primers and probes) in order to determine copy numbers and thereby determine the amount of GE materials. Single copy reference genes are generally preferred for absolute quantification of GE events by ddPCR. The suitability of four reference genes (HMG-I/Y, FatA(A), CruA and Ccfp) for absolute quantification of GE canola events by ddPCR was investigated. Variability was observed among the four reference genes in terms of copy numbers generated as well as consistency. Real-time quantitative PCR results were not affected by the use of single and two copy reference genes. However, ddPCR results were affected by the use of single vs. two copy reference genes. It is important to make an adjustment for two copy reference genes in order to obtain the expected quantification results with ddPCR.

P0167: Methods: Other Genome Methodology

Targeted Recombination to Increase Genetic Gain in Self-Pollinated Crops

Sushan Ru, University of Wisconsin-Madison, University of Minnesota, Madison, WI and Rex Bernardo, University of Minnesota, St. Paul, MN

Targeted recombination refers to the ability to have recombination at any genome position of interest. Targeted recombination can potentially be induced during mitosis through CRISPR technology, or it can be achieved by selecting for and pyramiding natural recombinations occur during meiosis. A previous study indicated that targeted recombination could double the rate of genetic gain for yield in maize, a cross-pollinated species for which hybrids are grown. To further investigate the potential of targeted recombination across different crops, we estimated prospective gains from targeted recombination in self-pollinated species. Previously published genotypic and phenotypic data on 21 biparental populations of barley, pea, soybean, and wheat were analyzed. Genomewide marker effects were estimated by ridge regression-best linear unbiased prediction. The performance of a doubled haploid with ideal single- or double-recombination on all chromosomes was used to calculate genetic gain. Our results showed that selecting for individuals with optimal single-recombination per chromosome could significantly ($P = 0.05$) increase gains by 25-274% compared to nontargeted recombination for all traits, except for plant height in barley. Prospective gains increased by 35-391% with optimal double-recombination per chromosome compared to nontargeted recombination. Overall, our study suggested much potential in targeted recombination as a breeding method in self-pollinated crops.

P0168: Methods: Other Genome Methodology

Statistical Inference of Phylogeny in Mixed-Ploidy Populations Using Genomic Variants

Han Qu[1], Yi Huang[1], Ruidong Li[2], Amy Litt[1] and Zhenyu Jia[1], (1)University of California Riverside, Riverside, CA, (2)University of California, Riverside, CA
Woody shrubs--manzanita (*Arctostaphylos* Adans.) is an ideal model organism as for advancing the knowledge of how the long and ongoing history of genetic exchange, frequent hybridization, restricted distribution, habitat and life history have promoted species diversification. Advanced sequencing techniques, such as Restriction-site Associated DNA (RAD) tags sequencing, allow us to find evidence at the whole genome molecular level rather than the conventional phenotype and limited genetic markers based phylogeny which has made research on a certain species a taxonomic quagmire. One of the major data challenges is the allele frequency ambiguity due to the combination of polyploidy (tetraploidy) in which more than two alleles may be involved. We develop an advanced statistical method for inference of the genotypes of polyploid individuals based on genotype data from a diploid population. The inferred genotypes for the polyploid individuals can be combined with the genotypes for the diploid individuals to calculate the genetic distances among mix-ploidy individuals using a novel approach, facilitating further genetic structure and phylogeny study. We demonstrate the method using intensively simulated data, and will test it on a RAD-seq dataset of manzanita. The project will provide software solutions to dealing with polyploidy in analyses of genetic diversity, a pervasive problem in plant biology.

P0169: Methods: Other Genome Methodology

Application of DNA Metabarcoding for Scattered Pollens in the Urban Atmosphere of Tokyo

Keisuke Tanaka[1], Yu Hamaguchi[1] and Shunsuke Yajima[2], (1)NODAI Genome Research Center, Tokyo University of Agriculture, Tokyo, Japan, (2)Department of Bioscience, Tokyo University of Agriculture, Tokyo, Japan
One of four Japanese people is troubled by pollinosis. Woodland maintenance by tree felling and artificial pruning in various local forest areas are performed as one of the reduction strategies of this national malady. Here, we think that a new woodland maintenance strategy that lead to effective tree felling and artificial pruning is constructed if a forest area of a pollen source scattered in the urban area was identified well. This study therefore performs environmental DNA metabarcoding analysis of pollen scattered in atmosphere and aims at the construction of a new environmental monitoring system. In this presentation, we report our technical methodology and a metabarcoding analysis result of the gathered pollen from the urban atmosphere of Tokyo.
A barcode region in *rbc*L gene of plastid genome was amplified from extracted DNA from pollen that sampled every day. Read data of the amplicon was outputted using high-throughput DNA sequencer. After data preprocessing and sequence clustering, organism species was explored by homology search on BLAST program.
Woody plants such as ginkgo and pine were detected at high frequency. Some polymorphic patterns were detected within sequences data estimated as the same species. Thus, we suggested the possibility of a local estimate by environmental DNA metabarcoding analysis of pollen scattered in the urban atmosphere.

P0170: Methods: Other Genome Methodology

Genomic Open-Source Breeding Informatics Initiative Data Extract and Decision Support Tools

Star Yanxin Gao[1], Philip Glaser[1], Joshua Lamos-Sweeney[1], Yaw A. Nti-Addae[1], Kevin Palis[1], Syed Raza[1], Angel Villahoz-Baleta[1], Deb Weigand[1], Elizabeth Jones[1], Edward S. Buckler[2], Jean-Luc Jannink[3], Lukas Mueller[4], Kelly Robbins[5], Mark E Sorrells[5], Qi Sun[5] and Susan McCouch[5], (1)Genomic Open-Source Breeding Informatics Initiative Project, Ithaca, NY, (2)USDA-ARS-Cornell University, Ithaca, NY, (3)USDA-ARS / Cornell University, Ithaca, NY, (4)Boyce Thompson Institute, Ithaca, NY, (5)Cornell University, Ithaca, NY
Genomic Open-source Breeding Informatics Initiative (GOBII) is a global community of multidisciplinary teams at Cornell University, USDA, the Boyce Thompson Institute, and International Maize and Wheat Improvement Center (CIMMYT), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), and International Rice Research Institute (IRRI), James Hutton Institute (JHI) and Diversity Array Technology (DArT). Our mission is to transform breeding by developing genomic data management systems, putting the tools in place to enable the implementation of genomic and marker assisted selection as part of routine breeding programs for staple crops in the developing world. The initial main crops are: rice, wheat, maize, sorghum and chickpea in South Asia and Sub-Saharan Africa expanding to many crops worldwide.
The genomic data management systems include databases, loader, extractor, analysis pipelines and decision support tools for plant breeders. Key capabilities to be presented include genotyping data management, quality control, pedigree verification, marker-assisted backcrossing, and genomic selection pipeline.
Complete list of people on the project can be found http://gobiiproject.org/

P0171: Methods: Other Genome Methodology

Topics on Genomic Selection in Breeding Context

Jialu Wei, Tingting Guo, Jinyu Wang, Xianran Li and Jianming Yu, Iowa State University, Ames, IA
Genomic selection(GS) provides a tool to increase the genetic gain per unit time with shorter breeding cycle. Genotypic and phenotypic data of training population are combined to predict the breeding values of individuals in testing population. GS has become an active research area these years and been facilitated a lot especially by recent advances in third-generation sequencing. Compared with viewing GS as a highly technical model building process, it is more interesting and meaningful to put it in breeding context, which is the initial objective. There are many topics studied on GS to solve problems and achieve higher genetic gains in breeding process. Here we will discuss three of them: 1) making genomic selection under genotype by environment interactions; 2) maintaining genetic diversity with genomic selection; 3) combing high-throughput phenotyping in genomic selection. Although solutions may not be available to some problems yet, enriching our understanding about these topics can be beneficial for further study.

P0172: Methods: Other Genome Methodology

Evaluation of Multiple Approaches to Identify Genome-Wide Polymorphisms in Closely Related Genotypes of Sweet Cherry (*Prunus avium* L.)

Seanna Hewitt, Molecular Plant Sciences, Washington State University, Pullman, WA

Identification of genetic polymorphisms and subsequent development of molecular markers is important for marker assisted breeding of superior varieties of economically important species. Sweet cherry (*Prunus avium L.*) is an economically important, non-climacteric tree fruit crop in the Rosaceae family and has undergone a genetic bottleneck due to breeding—This has resulted in limited genetic diversity in the sweet cherry germplasm that is utilized for breeding new varieties. Therefore, it is critical to recognize the best platforms for identifying genome-wide polymorphisms that can help identify and consequently preserve the diversity in a genetically constrained species. For the identification of polymorphisms in five closely related genotypes of sweet cherry, a gel-based approach (TRAP), modified reduced representation sequencing (TRAPseq), a 6k cherry SNParray, and whole genome sequencing (WGS) approaches were evaluated in the identification of genome-wide polymorphisms in sweet cherry cultivars. All platforms facilitated detection of polymorphisms among the genotypes with variable efficiency. In assessing multiple SNP detection platforms, that study has demonstrated that a combination of appropriate approaches is necessary for efficient polymorphism identification, especially between closely related cultivars of a species. The information generated in this study provides a valuable resource for future genetic and genomic studies in sweet cherry, and the insights gained from the evaluation of multiple approaches can be utilized for other closely related species with limited genetic diversity in the germplasm.

P0173: Methods: Other Genome Methodology

Optimizing Trait Predictability in Hybrid Rice using Superior Prediction Models and Selective Omics Datasets

Shibo Wang[1], Julong Wei[1], Ruidong Li[2], Han Qu[1], Weibo Xie[1] and Zhenyu Jia[1], (1)University of California Riverside, Riverside, CA, (2)University of California, Riverside, CA

Hybrid breeding has dramatically boosted yield and stability in rice. Genomic prediction further benefits rice breeding by increasing selection intensity and accelerating the breeding cycle. With the rapid advancement of technology, other omics data, such as metabolomic data and transcriptomic data, are readily available for predicting genetic values (or breeding values) for the agronomically important traits. In the current study, we searched for the best prediction strategy for four traits (yield, 1000 grain weight, number of grains per panicle and number of tillers per plant) of hybrid rice by evaluating all possible combinations of omics datasets with different prediction methods. We conclude that, in rice, the predictions using the combination of genomic and metabolomics data generally produce better results than single-omics predictions or predictions based on other combined omics data. Inclusion of transcriptomic data does not improve predictability possibly because transcriptome does not provide more information for the trait than the sum of genome and metabolome; rather, the computational complexity is substantially increased if transcriptomic data is included in the models. Best linear unbiased prediction (BLUP) appear to be the most efficient prediction method compared to the other commonly used approaches, including LASSO, SSVS, SVM-RBF, SVP-POLY and PLS. Our study has provided a guideline for selection of hybrid rice in terms of which types of omics datasets and which method should be used to achieve higher trait predictability.

P0174: Methods: Other Genome Methodology

Optimized Methods for High Molecular Weight Genomic DNA Isolation

Erwan Denis, Sophie Sanchez, Ghislaine Magdelenat, Corinne Cruaud, Arnaud Lemainque, Jean-Marc Aury and Valerie Barbe, CEA - Genoscope, Evry, France

Isolation of high quality ultra-High Molecular Weight (uHMW) DNA is a prerequisite for new technologies as long read sequencing (e.g. Pacific Biosciences, Oxford Nanopore Technology and 10x Genomics) and also for optical mapping (e.g. Bionano Genomics). Indeed, traditional DNA extraction methods lead generally to 50kb long DNA, generally called HMW DNA, but this is not enough for some technologies. Moreover, it often contains too much Low Molecular Weight DNA that could spoil the results of those new technologies. The need to extend HMW to uHMW appears to be crucial for long read sequencing and essential for optical mapping.

On another hand, DNA isolation can be very tricky depending on the organism of interest. For example, for plants this is very complicated due to the cell wall and secondary metabolites that could affect DNA yield and quality.

The classical uHMW DNA isolation method is based on the isolation of cells or nuclei in agarose gel plugs. Nowadays, despite it is time consuming and requires handling expertise, this method is the only one recommended for optical mapping. Here we present the development of an alternative method to purify and isolate high quality uHMW from different organisms (yeast and plants).

P0175: Methods: Other Genome Methodology

Updates to Farm Animals Annotation in Ensembl

Thibaut Hourlier, Konstantinos Billis, Carlos García Girón, Leanne Haggerty, Denye Ogeh, Osagie Izuogu, Fergal J. Martin, Bronwen Aken and Paul Flicek, European Molecular Biology Laboratory - EBI, Cambridge, United Kingdom

As a result of improvements to the scalability and automation our pipelines, we can annotate more species, in a shorter amount of time, while still producing high quality annotations. We will present two genome annotations generated with the Ensembl Gene Annotation System using publicly available data sets.

We updated the pig annotation using the improved pig reference assembly, Sscrofa11.1 (GCA_000003025.6), produced by the Swine Genome Sequencing Consortium. We annotated the latest goat assembly, ARS1 (GCA_001704415.1), produced by USDA-ARS.

Both assemblies have a high contig and scaffold N50, greater than 25Mb, which means longer sequences and a better representation of repeat regions. Two thirds of the models are supported by species-specific data coming from cDNAs and transcriptomic data such as Illumina RNA-Seq and PacBio Iso-Seq. We used transcriptomic data from different tissue samples to annotate untranslated regions (UTRs), alternate transcript isoforms. Furthermore, for each released annotation we update gene-trees, orthologues, multiple whole genome alignments, and cross-references to external databases.

The pig data are available in Ensembl (version e91) and are accessible through our website, www.ensembl.org, our REST API, http://rest.ensembl.org, BioMart (http://www.ensembl.org/biomart) and our public MySQL server, ensembldb.ensembl.org. Custom data can be displayed in the Ensembl browser. The goat annotation is expected for release 92 of Ensembl.
We support upload of data in multiple file format such as BAM or GFF. For groups who wish to share their datasets and view them alongside Ensembl data, we have developed a TrackHubRegistry (http://trackhubregistry.org/) to enable discovery of publicly accessible track data hubs.

P0176: Methods: Other Genome Methodology
Optical Mapping with the Saphyr System on Plant Genomes
Erwan Denis, Caroline Belser, Wahiba Berrabah, Ghislaine Magdelenat, Corinne Cruaud, Arnaud Lemainque, Jean Marc Aury and Valerie Barbe, CEA - Genoscope, Evry, France
Optical mapping is a technique consisting of the assembly of digested ultra-High Molecular Weight (uHMW) DNA molecules in order to construct restriction maps improving genome assemblies.
Bionano Genomics technology combines proprietary NanoChannel arrays with optical mapping to image extremely long uHMW DNA. The release of the new device Saphyr from Bionano Genomics is a real progress for optical mapping, especially for big genomes due to automation and yield of data obtained.
On Musa and Brassica species, we have generated more than 200Gb data from molecules above 150kb and obtained BspQI maps N50 of 6.5Mb and 2.25Mb respectively. Furthermore, we will show hybrid scaffold results from our lab with the Saphyr system on those two different plant species.

P0177: Methods: Other Genome Methodology
Benchmarking High-Molecular-Weight DNA and Tissue Preservation Protocols at the Vertebrate Genome Lab
Jacquelyn Mountcastle, Bettina Haase, Sadye Paez, Erich D. Jarvis and Olivier Fedrigo, The Rockefeller University, New York, NY
At the Vertebrate Genome Laboratory (VGL) within the Rockefeller University, we aim to generate high-quality, phased, near chromosome-level, annotated, reference genome assemblies of all living vertebrates for the Vertebrate Genome Project (VGP). The VGP is an international and multidisciplinary project of the Genome 10K (G10K) consortium in collaboration with Bird 10K, Bat1K and other consortiums. The VGP has selected genomic technologies with long-range sequence information, including Pacific Biosciences long reads, 10X Genomics linked reads, Hi-C chromosomal linked reads, and Bionano Genomics optical maps, which were found necessary to generate more complete, contiguous, and haplotype phased genome assemblies. To be fully successful, these genomic technologies require high-molecular weight (HMW) DNA. Obtaining sufficient amounts (> 5ug) of HMW DNA (30-150 Kb) can often be challenging. The yield and quality of DNA isolation is not only influenced by isolation protocols but also affected by tissue types and preservation methods. Here we present a comparative study of HMW DNA isolation protocols for various sample types (muscle, liver, blood) across multiple preservation methods. We have identified multiple combinations yielding several micrograms of DNA with fragments greater than 50kb optimal for these genome sequencing platforms and high-quality reference genomes.

P0178: Methods: Other Genome Methodology
Recombination Generates Mitochondrial Genome Diversity in Wild and Domesticated Lettuce
Beth Rowan, University of California, Davis, Davis, CA, Alex Kozik, UC Davis, Davis, CA, Alan Christensen, University of Nebraska - Lincoln, Lincoln, NE, Dean Lavelle, Genome Center, University of California, Davis, CA, M. Eric Schranz, University of Amsterdam, Amsterdam, Netherlands and Richard Michelmore, Genome Center, University of California, Davis, Davis, CA
Plant mitochondrial genomes exhibit extensive variation in size, sequence arrangement, and repeat content, and yet coding sequences remain highly conserved. The mechanisms driving these differences remain unclear, but recombinational repair processes potentially have an important role. Many plant mitochondrial genomes have one or more pairs of large repeats that can act as sites for inter- or intramolecular recombination, leading to several alternative genomic arrangements. Most mitochondrial genome assemblies have been generated using methods that were unable to capture the complete spectrum of isoforms, leading to an incomplete inference of recombinational activity. We used a combination of long-read (PacBio) and short-read (Illumina) sequencing data to produce mitochondrial genome assemblies and characterize isoforms for wild (*Lactuca saligna*) and domesticated (*Lactuca sativa*) lettuce species. We also performed a microscopic analysis of the physical molecules of mtDNA. We found that both species had three pairs of large repeats (> 3 kb) and one smaller repeat (< 2 kb) that were recombinationally active, but the isomeric complexity and stoichiometry of the resulting isoforms differed immensely between species. The length and composition of the small repeat differed between *L. saligna* and *L. sativa,* and this was associated with differences in recombination frequency. Like other plants, we found most physical molecules of lettuce mtDNA in subgenomic circular, simple linear, branched linear, and complex multigenomic forms. This shows that the use of long-read sequencing technology combined with the microscopic analysis of mtDNA molecules provides insight into recombinational mechanisms that drive structural and physical diversity in plant mitochondrial genomes.

P0179: Methods: Other Genome Methodology
Full Spectrum Genome Analysis with Linked-Read Sequencing
Shamoni Maheshwari[1], Andrew Wei Xu[2], Sarah Garcia[2], Stephen Williams[3], Claudia Catalanotti[2], Michael Schnall-Levin[2], Patrick Marks[2] and Deanna M. Church[2], (1)10x Genomics, Inc, Plesanton, CA, (2)10x Genomics Inc., Pleasanton, CA, (3)10x Genomics, Inc
Highly accurate and affordable short-read based methods fail to give a complete picture of a genome and are limited by the lack of long-range information. They struggle to identify large, balanced structural events, cannot access repetitive regions of the genome, and fail to resolve diploid genomes into two haplotypes. Here we describe an approach that retains long range information while harnessing the power of short

reads. The 10x Genomics Chromium™ Genome solution utilizes haplotype-level dilution of high molecular weight DNA molecules into >1 million barcoded partitions to create a novel data type referred to as 'Linked-Reads' (LR). This approach enables high-resolution genome analysis with minimal DNA input (~1 ng).
Our reference-based pipeline, Long RangerTM , leverages the unique properties of LRs to call a broader range of variant types. Barcoding of individual DNA molecules enables parsing of heterozygous variants into distinct haplotypes, providing increased power for variant calling, particularly for balanced events and single exon deletions and duplications. Using a well-characterized human truth sample (NA12878), we demonstrate the ability to call and haplotype challenging variants including balanced structural variations, heterozygous deletions, and SNPs on distinct paralogous loci.
Our Supernova™ Assembler takes advantage of Linked-Reads to perform *de novo* diploid assembly. We demonstrate the performance of Supernova on diverse plant and animal species. Recent updates in the Supernova algorithms allow for assembling small genomes (<1Mb) using Linked-Reads. To demonstrate these new features, we constructed a diploid assembly for the ~500Mb Flame Grape genome.

P0180: Methods: Other Genome Methodology
Efficient Hybrid Breeding Based on the Modification of Meiosis using Virus-Induced-Gene-Silencing
Vanesa Calvo, WUR Wageningen University, Wageningen, Netherlands
Vanesa Calvo, Linus Hohenwarter, Cris Wijnen, Hans de Jong, Bastiaan de Snoo, Arp Schnittger and Erik Wijnker.
Wageningen University, Wageningen, The Netherlands. Hamburg University, Hamburg, Germany.
Hybrid production in traditional breeding programmes is usually a lengthy process. Obtaining new inbred lines or finding favourable parental combinations requires a great investment of resources. To overcome these issues, we have developed a new breeding approach to produce hybrids more efficiently. First, it allows to select and fix any unknown hybrid genotype. Secondly, we can also obtain near-full hybrids that can perform either as the initial hybrid or potentially better. Our new breeding technique is based on the reduction of 80% of meiotic crossovers, by silencing MSH5 directly in the hybrid, using Virus-Induced-Gene-Silencing (VIGS). The major advantages of this new application are: i) absence of stable transgenes to modify gene expression ii) rapid generation of hybrids (only 3 generations from the initial hybrid) and iii) the production of new parental lines that are either chromosome substitution lines or low-recombinant lines (1 or 2 recombination events). This efficient hybrid breeding approach based on the modification of meiosis brings a new insight to hybrid production and evaluates hybrid performance by comparing full hybrids with rapidly obtained near-full hybrids.

P0181: Methods: Other Genome Methodology
Exploring Heteroplasmy As the Basis for Maternally-Transmitted Cold Tolerance in Cucumber
Madeline W. Olberg, University of Wisconsin, Madison, WI and Michael J. Havey, USDA-ARS and University of Wisconsin, Madison, WI
Cucumber is a warm-season crop that can be severely damaged by short periods of cold temperatures. Growers would benefit from cold tolerant cucumbers by preventing crop loss in inclement weather as well as by allowing for earlier planting and harvest, thus avoiding heavy late-season disease pressure. A maternally-inherited cold tolerance exists in the heirloom variety, 'Chipper.' In cucumber, the chloroplast, mitochondrial, and nuclear genomes are maternally, paternally, and biparentally transmitted, respectively, indicating that this cold tolerance may be conditioned by the chloroplast genome. Heteroplasmic variants could condition a nuclear response in 'Chipper' to quickly respond to sudden cold stress. Chloroplast DNAs were sequenced from doubled haploids (DH) of cold tolerant 'Chipper' and susceptible 'Straight 8' and '9930.' Heteroplasmy for SNPs between cold tolerant and susceptible cucumbers were identified. Variation will be confirmed by RNA-seq and targeted proteome analysis of the reciprocal hybrids of these DH lines. Identification of the genetic basis of cold tolerance in 'Chipper' should provide potential targets of selection for cold tolerance in cucumber, as well as other warm-season crops.

P0182: Methods: Other Genome Methodology
Pattern of Gene Expression during Shift in Ploidy in *Funaria hygrometrica*
Nasim Rahmatpour, University of Connecticut, Storrs, CT
Signatures of past WGD are detected in all taxa of bryophytes for which genomic or transcriptomic data are available. Such genome doubling may arise from allopolyploidy or autopolyploidy. Consequences of autopolyploidy on gene expression are not well studied in plants, and to date, no studies have examined the effect of whole genome duplication in bryophytes. Here, we examine the effect of whole genome duplication on gene expression profiles by comparing the transcriptome of haploid gametophytes and first generation artificial isogenic diploid gametophytes that were generated via apospory. Specifically**,** we are testing whether the first generation of autopolyploid plants exhibit a different expression profile compared to their progenitors. The top portion of pre-meiotic, immature, spear-shaped sporophytes were sampled. Transcriptomes were provided from three technical replicates of haploid and diploid aposporous gametophyte tissues. The Illumina paired-end reads were assembled into transcripts via TRINITY and clustered via USEARCH to generate a single reference of 53,283 and 64,246 unique transcripts in the haploid and diploid gametophyte, respectively. To compare the gene expression among transcriptomes of haploid and diploid gametophytes, we used GFOLD for all replicates of same data set and assigned a reliable fold change for each gene. The DESeq2 package (R Bioconductor) followed for advanced normalization and analysis of differential expression to examine evidence for duplication.
Submission email address: nasim.rahmatpour@uconn.edu

P0183: Methods: Other Genome Methodology
Genomic Structural Variation across Five Continental Populations of *Drosophila melanogaster*
Evan Michael Long[1], Carrie R. Evans[2] and Joshua A. Udall[2], (1)Brigham Young University, Orem, UT, (2)Brigham Young University, Provo, UT
Chromosomal structure variations (SV) including insertions, deletions, inversions, and translocations occur within the genome and can have a significant effect on an organism's phenotype. Some of these effects are caused by structural variations containing genes. Modern sequencing

using short reads makes the detection of large structural variations (> 1kb) very difficult. Large structural variations represent a significant amount of the genetic diversity within a population.
We used a global sampling of *Drosophila melanogaster* (Ithaca, Zimbabwe, Beijing, Tasmania, and Netherlands) to represent diverse populations. We used optical genome mapping technology to identify structural variations in these genomes. Because the average read length used for optical mapping is >150kb, these maps facilitate the identification of chromosomal SVs of greater size and with more clarity than sequencing based approaches. We have also produced long read sequence data to validate the identified structural variations.
We found unique and shared structural variations in each of the five strains. The frequency and relationship of the structural variations correlate with previously performed single nucleotide polymorphism (SNP) studies. Structural variations accounted for a much larger difference in number of base pairs between strains than SNPs. Large structural variations make up a significant portion of variation among populations.

P0184: Methods: Sequencing
Leveraging Public Transcriptome Data for Enhanced qPCR Probe Design
Sumyya Waliullah, Tree Fruit Research & Extension Center, Washington State University, Wenatchee, WA, Loren A Honaas, USDA-ARS Tree Fruit Research Lab, Wenatchee, WA and Lee A. Kalcsits, WSU Tree Fruit Research and Extension Center, Wenatchee, WA
Quantitative real-time PCR (qPCR) is the gold standard for targeted gene expression analysis and is a critical validation tool for RNA-seq experiments. And as plant genomes continue to become available for economically and scientifically important plants, opportunities to leverage these genomic resources for closely related plants (i.e. cultivars, species, ecotypes) arise. However, qPCR (and RNA-Seq) rely on prior knowledge of gene sequences, thus genetic differences between sequenced plants and closely related plants of interest can cause loss of fidelity in gene expression measurements at multiple steps in an experiment. To avoid this pitfall, we used publicly available transcriptome data from Honey Crisp apple to discover cultivar specific variants and guide our qPCR primer design strategy to study genes potentially involved in the physiological fruit disorder, bitter pit. Bitter pit is a calcium-related physiological disorder that typically manifests as corky, dark and depressed spots on the fruit surface which cause high post-harvest losses in apple. Towards predictive and diagnostic tests, knowledge of molecular mechanism underlying BP development are essential. To check BP related gene expression we chose genes from a previous report that are implicated in the disorder. Using BLAST we queried a de novo transcriptome assembly of Honey Crisp with coding sequences of Golden Delicious genes or cultivar specific clones. We used the Honey Crisp best hits to design qPCR primers with 100% success – all primer pairs produced target specific amplicons with high efficiency. Furthermore, we were able to learn about cultivar specific gene variants and, by examining alignments to discover polymorphisms that would impact primer binding, opened the door to qPCR assays that are variant specific or have cross-cultivar applicability. This work is possible with open-source software and freely available public data, necessitating no upfront cost.

P0185: Methods: Sequencing
NC State University Genomic Sciences Laboratory
David A. Baltzegar, Genomic Sciences Laboratory/ NC State University, Raleigh, NC
The Genomic Sciences Laboratory (GSL) at NC State University was established in 2000 to enable biological discovery by providing the necessary tools to conduct advanced genomics research. As a land-grant university facility, the GSL has an extensive focus in Agrigenomics, recently providing genome sequencing and genotyping data for Striped bass (*Morone saxatilis*), blueberry (*Vaccinium corymbosum),* blackberry *(Rubus sp.)*, sweet potato (*Ipomoea batatas*) and other emerging cultivars, as well as domestic food products (*e.g.*, metagenomic analysis of farmed sturgeon caviar). Administered by the NC State Office of Research, Innovation & Economic Development, the GSL operates as a solution center for NC State faculty, their collaborators, and the broader scientific community (both U.S. and International). The GSL is staffed by a facility Director and four full-time research technicians, with over 50 years combined experience in molecular biology and Next Generation Sequencing (NGS) approaches (e.g. RNAseq, *de novo* genome sequencing, and ddRADseq). We currently offer traditional Sanger DNA sequencing and microsatellite genotyping, as well as NGS services at highly competitive pricing to maximize discovery potential for our users. Illumina MiSeq, NextSeq 500, and HiSeq 2500 sequencers are housed in our facility, as well as a new, state of the art PacBio Sequel for long-read NGS sequencing. Additional services include RNA extraction from tissues, nucleic acid quality analysis by Agilent Tapestation, and Illumina and PacBio NGS library preparation. Along with a sister core facility providing bioinformatic services (NCSU Bioinformatics Consulting and Service Core), GSL services are available on a cost-recovery basis to all users.

P0186: Methods: Sequencing
Improved Workflows for Genome, Transcriptome and Metagenomic Sequencing
Chris L. Wright, Mark Band, Mark A. Mikel and Alvaro G. Hernandez, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL
New technological advancements in sequencing methodologies continue to enable genomic research that was unthinkable a few years ago. We are applying a combination of next-generating sequencing tools across a wide array of applications in plants, animals and metagenomic communities, including whole genome sequencing, genome resequencing and genotyping, characterization of gene expression, epigenomics, and metagenomics.
In this poster, we show how blending the power of the 10x Genomics Chromium system, Illumina sequencing platforms including the NovaSeq, and the Oxford Nanopore GridION facilitate a thorough characterization of genomes and their regulation in a cost-effective manner. We specifically discuss how we apply these new tools to crop studies, as well as animal, environmental, and nutritional sciences projects. We also present a new pipeline for shotgun metagenomic DNA samples that allows for a more unbiased view of the microbial community present in the sample and provides for easier sequence analysis with minimal need for computational resources.
Furthermore, we present a novel microbial profiling system that allows simultaneous interrogation of 16S, 18S, ITS, archaeal, and functional genes. This workflow produces an amplicon pool ready to sequence on the Illumina systems at a cost significantly reduced compared to whole metagenomic sequencing, while still allowing for classification of all organisms and desired functional genes within a microbial sample. It can also be applied to any plant or animal species of interest to study hundreds of desired loci.

P0187: Methods: Sequencing

gVolante: Multi-Faceted Evaluation of *de novo* Genome Assemblies

Osamu Nishimura, Yuichiro Hara and **Shigehiro Kuraku**, RIKEN, Kobe, Japan

Along with the explosive increase of accessibility to various genome sequencing technologies, the demand for assessing the quality of their products has been multiplied. To this end, metrics based on sequence lengths, such as N50, have become a standard, but they evaluate only one aspect of assembly quality. Conversely, analyzing the coverage of pre-selected single-copy orthologs provides essential content-based quality assessment. Here, we introduce a web server **gVolante** (https://gvolante.riken.jp/) which provides an online interface for (i) on-demand completeness assessment of genome assemblies with the program pipelines CEGMA and BUSCO and (ii) browsing pre-computed completeness scores for publicly available data (Nishimura et al., 2017. *Bioinformatics*. doi.org/10.1093/bioinformatics/btx445). In gVolante, one can select our original reference ortholog set, CVG (Core Vertebrate Genes) previously introduced to increase the assessment accuracy for vertebrates (Hara et al., 2015. *BMC Genomics*. 16:977). Completeness assessment performed on the gVolante web server reports both coverage of reference orthologs and statistics regarding sequence lengths (e.g. N50 scaffold length), allowing multi-faceted quality control. Using gVolante, one can compare the quality of original assemblies between their multiple versions (obtained through program choice and parameter tweaking, for example) and evaluate them in comparison to the scores of public resources. This presentation will cover some updates of the web server functions to accommodate to diversifying sequencing targets and strategies.

P0188: Methods: Sequencing

Integrated Search of Public NGS Data for Non-Canonical Model Organisms

Takeru Nakazato, Tazro Ohta and Hidemasa Bono, Database Center for Life Science, Mishima, Japan

High-throughput sequencing, also called next-generation sequencing (NGS), is widely used for omics analysis with non-classical model organisms, and its numerous sequence read data have been archived in public database, the sequence read archive (SRA) with their experimental information as metadata by DDBJ, EBI and NCBI. As of October 2017, 13.9 peta bp of reads are archived in about 86,000 projects.

We categorized SRA data by study types (e. g. whole genome study, transcriptomics, and metagenomics), sequencing platforms, and species or cell lines of samples. To visualize these information, we developed a web-service called DBCLS SRA (http://sra.dbcls.jp/).

Recently, project information archived in SRA has moved to BioProject database, and sample information has also moved to BioSample databases. In addition, transcriptome data with NGS have archived in not only SRA but also GEO as gene expression databases. We therefore connected corresponding submissions in various databases.

For plant and animal researchers, DBCLS SRA provides the NGS data list of sample species or strains under same taxonomy level such as species, genus, families in one search. For example, they can retrieve Oryza sativa indica or japonica strain data by searching Oryza sativa. Users also list up other Oryza species data as a parent taxonomy.

P0189: Methods: Sequencing

Improved Methods for Next Generation Sequencing Library Cleanup and Size Selection

Charles Cowles[1], Douglas Horejsh[1], Curtis Knox[1], Mark Denhart[1], Adam Blatter[2], **Tim Deschaines**[2] and Marjeta Urh[1], (1)Promega Corporation, Madison, WI, (2)Promega Corporation, Madison, WI, Madison, WI

NGS libraries require high quality nucleic acid inputs of varying quantity, concentration, and size depending on the library preparation methods and sequencing platforms used. Most methods use a magnetic bead-based chemistry throughout the overall protocol, with usage falling into two categories of function: 1) Sample clean up to remove unwanted components (adaptors, primers, dNTPs, enzymes, etc.), and 2) Size selection to remove nucleic acid fragments or library molecules outside of an optimal size range for downstream sequencing. Although commercially available beads are used extensively in NGS library prep methods, there are performance issues that can be improved upon, including significant DNA loss, poor reproducibility, difficulty in pipetting due to viscosity of chemistry, and retention of undesired high molecular weight DNA.

Comparisons between commercial size-selection beads and a newly developed chemistry were performed to assess library yield, reproducibility, size-selection precision, and sequencing results. WGS was performed using the NEBNext UltraII library kit on a range of DNA inputs from 1ng - 1µg. Size selection steps were executed in parallel with both size-selection chemistries, followed by Illumina sequencing. Additionally, measurements were conducted to determine viscosity and bead response time between chemistries as these parameters can significantly affect reproducibility.

DNA yields increased ≥20% per sample cleanup step cycle with the newly developed chemistry, and substantial reduction of undesired HMW DNA from libraries was achieved. Viscosity and bead response time are both reduced at least 5-fold, resulting in greater reproducibility and 2 to 4-fold reduction in library yield standard deviation. Sequencing results demonstrating duplication rates, coverage uniformity and reproducibility will be presented.

Improved DNA recovery from NGS library size-selection can enable working with smaller starting sample inputs and reductions in PCR cycling, resulting in fewer duplicate reads, greater coverage uniformity, and higher confidence in mutation calling. Reduced variability and better removal of HMW fragments in sample preparation can minimize costly sequencing failures and wasted reads.

P0190: Methods: Sequencing

State of the Art and Comparison of Long Reads Technologies

Celine Vandecasteele, GeT-Plage Core Facility - INRA France, Castanet tolosan, France and Claire Kuchly, GeT-PlaGe Core Facility - INRA France, Castanet Tolosan cedex, France

Thanks to its experience on short reads sequencing using the Illumina technologies, the GeTPlaGe core facility began to evaluate and use long read technologies since the beginning of 2015 : Pacific BioSciences RSII, Oxford Nanopore Technology MinION and 10X Genomics

Chromium. Genomic issues such as complex genome assembly, structural variant discovery or phasing can be addressed by those long read technologies.
As DNA quality is the most important requirement to obtain an efficient sequencing, sample requirements for each technology, and quality controls performed on GeT-PlaGe will be detailed. For all the technology presented, DNA sample needs high quality and purity. For ONT MinION and 10X Genomics Chromium, there ads length have theoretically no limits compare to PacBio RSII (max around 50kb) but the input DNA size is the key for all of them. The amont of DNA required for sequencing can be huge and challenging to obtain. The DNA quality is the corner stone of a good bioinformatics analysis particulary for assemblies.
We are presenting current projects concerning de novo assembly results obtained using multiple Long Read technologies, for several genomes (bacteria, fungus, tomato and fish).

P0191: Methods: Sequencing

CleanPlex's Simple Workflow Produces NGS Libraries with Low Background and High Variant Detection Accuracy

Lucie S. Lee, Yang Lily Liu, Jeffrey Liu, Guoying Liu, Jacey Li Zhang and **Zhitong Liu**, Paragon Genomics, Hayward, CA

Next- generation sequencing (NGS) technologies have accelerated research efforts in the fields of biomedical sciences, clinical diagnostics, environmental genomics, and plant and animal agrigenomics. Especially with multiplex PCR (mPCR), scientists are capable of utilizing NGS to detect and analyze gene sequences in humans, animals, microorganisms, and plants for species authentication, de novo sequencing, and variant identification. Amplicon- based targeted library preparation methods, utilizing mPCR, are preferred by many as the more time and cost-efficient method over hybrid-capture methods.
However, there are common drawbacks with amplicon-based methods as well. Multiplex PCR has inherent difficulties such as high background and poor uniformity. We developed a high-performing 2.5hr, one tube, amplicon-based library preparation solution for NGS that addresses and resolves these drawbacks. We present results from experiments demonstrating the robust performance of the Paragon Genomics CleanPlex System that yield libraries with low background, no apparent GC bias, and >95% uniformity with as little as 1 ng of DNA. The system has a limit of detection below 1% and high variant detection accuracy, as well as high mapping and on-target rates over a wide range of target numbers and input DNA amounts.

P0192: Methods: Sequencing

Novel Method for the Preparation of Target-Enriched Plant Libraries for Genotyping By Next Generation Sequencing.

Cynthia Hendrickson[1], Kruti Patel[1], Amy Emerman[1], Charles Elfe[1], Scott Adams[1], Bjoern Textor[2], Andrew Barry[3], Sarah Bowman[1] and Evan Mauceli[1], (1)Directed Genomics, Ipswich, MA, (2)New England Biolabs, GmbH, Frankfurt am Main, Germany, (3)New England Biolabs, Ipswich, MA

Genotyping by next generation sequencing has significant advantages over traditional genotyping methods, including the ability to interrogate thousands of markers at high resolution while discovering newly introduced SNPs that may contribute to phenotypic changes. Plant genotyping presents additional challenges in target enrichment design, library preparation, and data analysis, arising from incomplete or absent reference genomes, repetitive sequences, and the presence of enzyme-inhibiting contaminants in DNA samples.
To provide a simple and inexpensive method for NGS-based plant genotyping, we modified the NEBNext Direct target enrichment protocol to decrease the cost per reaction and tested it with a variety of plant species. In the NEBNext Direct protocol, DNA is enzymatically nicked, denatured, and enriched through hybridization-based capture, followed by enzymatic removal of non-targeted sequences and conversion into Illumina-compatible libraries. This approach resulted in 90-99% of the sequencing inserts mapping to targeted regions. Of note, blocking repetitive DNA regions during hybridization was not required to achieve this level of specificity. No difference in specificity was observed between species with a reference genome available for bait design versus species with only amplicon sequences available. In addition, libraries prepared from CTAB-purified DNA demonstrated only a small decrease in conversion compared to libraries prepared with higher quality, column-purified DNA. Thus, the modified NEBNext Direct protocol provides a cost-effective and robust solution for NGS-based plant genotyping that is compatible with contaminants present in CTAB-purified DNA and can be applied to species with little genomic information available.

P0193: Methods: Sequencing

A Novel and Efficient Approach for Phasing Highly Heterozygous Plant Genomes

Tingting Zhu[1], Le Wang[1], Frank M. You[2], Juan C. Rodriguez[1], Karin R. Deal[1], Limin Chen[1], Jie Li[3], Sandeep Chakraborty[1], Bipin Balan[1], Cai-Zhong Jiang[1,4], Patrick J. Brown[1], Charles A. Leslie[5], Mallikarjuna Aradhya[6], Abhaya M. Dandekar[1], Daniel A. Kluepfel[4], Jan Dvorak[1] and Ming-Cheng Luo[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada, (3)Genome Center, University of California, Davis, Davis, CA, (4)USDA-ARS Crops Pathology and Genetic Research Unit, Davis, CA, (5)Department of Plant Science, University of California, Davis, Davis, CA, (6)USDA National Clonal Germplasm Repository, Davis, CA

Assembling highly heterozygous plant genomes from short sequence reads is challenging due to difficulty in recovering the different haplotypes. Standard assembly protocols tend to collapse homozygous regions and report heterozygous regions as alternative contigs; such multiple assemblies are hard to resolve leading to fragmented assemblies larger than the expected size. We devised a novel method that overcomes genome heterozygosity by assembling two haploid genomes of an interspecific hybrid. Here we report the *de novo* assembly of two haploid genomes in interspecific hybrid MS1-56 (*Juglans regia* cv. Serr × *Juglans microcarpa*). We used a combination of BioNano genome (BNG) mapping, PacBio single-molecule real-time (SMRT) and Illumina sequencing technologies along with standard and custom designed assembly protocols to achieve complete assembly of two haploid genomes (*J. regia* and *J. microcarpa*) comprising the genome of hybrid MS1-56. By coupling SMRT sequencing and BNG mapping technologies, we were able to generate a 1.07 Gb highly contiguous assembly, with a contig N50 size of 8.0 Mb and a scaffold N50 size of 34.8 Mb. We also constructed BNG maps for both parental species of MS1-56 and successfully partitioned the two haplotypes from the sequence assembly of MS1-56, i.e. 529 Mb for *J. regia* 'Serr' and 538 Mb for *J. microcarpa*, respectively. We then applied the genetic map of *J. regia* cv. Chandler onto each assembled genome, resulting in 532 Mb scaffolds

in *J. regia* 'Serr' and 524 Mb scaffolds in *J. microcarpa* anchored onto 16 chromosomes in each genome, of which 12 and 14 chromosomes in *J. regia* 'Serr' and *J. microcarpa*, respectively, were able to be resolved into single scaffolds. To date, this work presents the most contiguous and complete genome assembly of a highly heterozygous plant species. It should also be noted that high-quality haplotype genomes for both parental species were generated from a single sequencing of one hybrid offspring.

P0194: Methods: Sequencing

Ultra-High-Multiplex and Customizable Targeted Sequencing Panels across Plant Species

Aren Ewing[1], Lifeng Lin[2], Shu Wang[2], Antonio Martinez-Alcantara[2], Heinz Breu[2], Manimozhi Manivannan[2] and Fiona Hyland[2], (1)Thermo Fisher, South San Francisco, CA, (2)Thermo Fisher Scientific, South San Francisco, CA

Plant and animal genomics research increasingly relies upon our ability to quickly and accurately generate genotype data on a large scale. Marker-assisted selection (MAS) requires the interrogation of a large number of genetic markers in parallel; phylogenetic and molecular evolution studies benefit from the sequencing of signature elements across both known and unknown species; adulterant detection in plant extracts calls for a highly sensitive method that is reliable, easy-to-use and capable of detecting miniscule amounts of sample.

In the last few years, highly accurate and uniform enrichment of targeted genomic elements, coupled with next-generation sequencing (NGS), is emerging as a unified solution for many of the aforementioned challenges. It provides the throughput and accuracy without the burden of whole genome sequencing that is difficult, expensive and complex for many plant species. However, some unique challenges exist in the design of targeted sequencing panels in plant species, including resolving closely-related species, designing assays to amplify a large number of species, designing massively multiplex assays, avoiding repetitive elements, and resolving primer cross-reactions.

Here, we describe targeted NGS solutions tailored to the needs of different aspects of plant genetic research, including a genotyping-by-sequencing (GBS) panel with ~2000 genetic markers; a species-identification panel; and an adulterant detection panel. We have developed a bioinformatics pipeline that can design customized panels that addresses the specificity and inclusivity concerns and generates the best quality, comprehensive, robust, multiplex assay design solutions with the minimum cost.

P0195: Methods: Sequencing

SeqSNP, a Massively Parallel Marker Screening Approach – High-Speed and High-Throughput Genomic Selection for Plants and Livestock

Joris Parmentier[1], Dr Berthold Fartmann[2], Silke Winkler[2], Dr. Samuel Arvidsson[2], Sabine Osterkamp[2], Dr Wolfgang Zimmermann[2] and Joe Don Heath[3], (1)LGC, Hoddesdon, United Kingdom, (2)LGC, Berlin, Germany, (3)NuGEN Technologies, San Carlos, CA

MAS (marker assisted selection) has seen massive advances in genetic gains linked to production and environmental tolerance for both plants and livestock. MAS uses markers in LD (linkage disequilibrium) with a QTL (quantitative trait loci) guided by phenotype, pedigree information and a relatively small number of SNP markers. Traits are, however, more often complex networks of genetic interactions requiring larger numbers of associated markers during selection and eventual trait fixation. Genomic selection (GS) is a method that uses high densities of markers (>~5K unique data-points per sample), interspersed throughout a genome, to increase association between migrating genetic material and phenotype, thus improving prediction accuracy for breeding and genotypic values.

For GS analysis to be of value, input tissue mass or DNA should be minimal along with quick turn-around time, returning data with high accuracy. Increasing allele number and concomitant increases in population sizes strains time, throughput and analysis resources for many to this kind of analysis. To remove these hurdles, LGC have developed SeqSNP, which tackles all of these problems with a complete sample to data service. Here we describe the utilisation of automation and Nugen's Allegro technology to genotype 50,000 alleles targets per sample in multiplex but also allows markers to be selected flexibly per project. We show a typical output using 96 tomato plants in a segregating population screened with 4744 alleles across the whole genome. We directly compared 297 marker subset tested with KASP showing high concordance of genotyping data between the two technologies.

P0196: Methods: Sequencing

Optical Maps Combined with Recent NGS Reach High Quality Genome Assembly in Plants

Sandrine Arribat[1], William Marande[1], Pierre Frasse[2], Pauline Heuillard[3], Olivier Bouchez[3], Mondher Bouzayen[2], Mohamed Zouine[2] and Helene Berges[1], (1)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (2)INPT-ENSAT, Toulouse, France, (3)INRA GeT Genomics Facility, Castanet-Tolosan, France

The French Plant Genomic Resources Center (CNRGV) is dedicated to the analysis of plant genome complexity. Indeed, among living organisms, plants display a high level of genome complexity due to their large size, variations in polyploidy levels and high percentage of repetitive elements. In a context of climate change, population growth and limited energy resources, increasing plant genomes knowledge is essential for a better understanding of mechanisms driving plant adaptation and evolution.

With the long reads sequencing technology, such as Pacific Bioscience that produces reads of several kilobases, the genome quality assembly was largely improved with a considerable contig number reduction and a better resolution of repeated sequences localization. More recently, a new technology of linked reads, named 10X Genomics, was used to create an additional scaffolding that gives a real input to the quality assembly. However, it remains challenging to obtain high quality assemblies at the genome scale.

At the CNRGV, in complement to different NGS technologies, we use whole genome optical maps to improve the genome assembly quality from several plant species. The Optical Mapping technology (ie. Irys system from Bionano Genomics) is based on direct visualization of high molecular weight DNA (from 150kb to 2Mb) labeled on specific sequences patterns. This hybrid scaffolding strategy (NGS and optical maps) could allow the production of genome resolved at the chromosome level. We will present a comparison of results obtained for the tomato genome with a combination of optical maps and either PacBio or 10X genomics sequencing technologies.

P0197: Methods: Sequencing

Application of Stellaris RNA FISH in Plants to Facilitate the Study of Gene Expression at the Single-Cell Resolution

Grace Lean[1], Wei Wu[2], Chloe Dowman[3], Yvain Chen[3], Jessica Kaplunov[3], Susan Duncan[4], Tjelvar Olsson[5], Matthew Hartley[5], Stefanie Rosa[6] and Caroline Dean[5], (1)LGC, Novato, CA, (2)LGC Genomics Ltd, Hoddesdon, United Kingdom, (3)LGC, Hoddesdon, United Kingdom, (4)Earlham Institute, Norwich, United Kingdom, (5)John Innes Centre, Norwich, United Kingdom, (6)University of Potsdam, Potsdam, Germany
Stellaris® RNA FISH (fluorescence *in situ* hybridization) using single molecule FISH technology (smFISH), is an RNA visualisation method that allows simultaneous detection, localisation, and quantification of individual RNA molecules in fixed cells and tissue. Stellaris RNA FISH probes are comprised of multiple oligonucleotides with unique sequences, each with a fluorescent label, that collectively bind along the same target transcript to produce a punctate signal. Post-hybridisation, samples are imaged using a fluorescent microscope.
Stellaris RNA FISH technology has been extensively applied to cells and tissues in model organisms ranging from human and mouse, to *C. elegans* and fungal pathogens. Research focus has spanned a wide range of topics including cancer, neuroscience and ageing. Until recently, plant biology has lagged behind other fields in the study of cell-to-cell variation of RNA due to the optical properties of plant cells.
Here we present the successful application of Stellaris RNA FISH in *Arabidopsis thaliana* meristem root tissue. Detection and image analysis of mRNAs from the housekeeping gene *PP2A* demonstrate, for the first time in plants, the quantification of single RNA transcripts. Subsequently, the mechanistic relationship at the cellular level between *FLOWERING LOCUS C* and *COOLAIR* has been explored, gaining a novel insight into cell-to-cell variability in expression levels.
The developed plant Stellaris protocol and associated open source image analysis algorithm will facilitate the application of this technique by other plant researchers. Stellaris RNA FISH technology follows a simple protocol, is inexpensive, platform-independent and offers same-day results, providing exciting opportunities for the research community.

P0198: Methods: Sequencing

Best Practices for Diploid Assembly of Complex Genomes using PacBio: A Case Study of Cascade Hops

Sarah B Kingan[1], David Rank[1], Kristin Mars[1], John Henning[2] and David Hendrix[3], (1)Pacific Biosciences, Menlo Park, CA, (2)USDA-ARS, Corvallis, OR, (3)Dept of Biochemistry and Biophysics, Oregon State University, Corvallis, OR
High quality reference genomes are essential resources for plant breeding, as well as functional and evolutionary studies. Outbred plants with highly repetitive genomes are challenging to assemble with standard short-read based methods. Diploid assembly with PacBio reads can generate assemblies with high contiguity and accurate haplotype resolution.
The common hop (*Humulus lupulus*) is a diecious plant in the Cannabaceae family. The cone-shaped flowers of the female plant are used to flavor and preserve beer. Cascade is the most widely used hop variety in American craft brewing. The genome of common hops is large (>2.5 Gb) and highly repetitive with individual plants displaying high levels of heterozygosity. We present a PacBio-only genome assembly with dramatic improvements in contiguity (contig N50 > 500kb) and gene content compared to earlier short-read assemblies for hops.
We describe the complete genome assembly workflow for this complex plant genome:

1. Sample preparation: extraction of high molecular weight DNA and production of large-insert SMRT bell libraries.
2. Genome assembly with FALCON-Unzip and polishing with arrow, with specific details of compute resources.
3. Analysis of resulting genome structure including raw read coverage and haplotype resolution.

P0199: Methods: Sequencing

Identification of Large Deletions in Mutant Soybean Lines from Genotyping-By-Sequencing Data

Marc-André Lemay[1], Davoud Torkamaneh[2], Guillem Rigaill[3], Adrian O. Stec[4], Robert M. Stupar[4] and Francois Belzile[1], (1)Laval University, Quebec, QC, Canada, (2)University of Guelph and Laval University, Quebec, QC, Canada, (3)IPS2 & LaMME, INRA & Université d'Évry Val d'Essonne, Orsay, France, (4)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
Genotyping-by-sequencing (GBS) allows efficient identification and genotyping of a large number of molecular markers by sequencing a constant subset of the genome captured through the use of a restriction enzyme. However, technical developments regarding this approach have been almost entirely focused on the discovery and genotyping of single nucleotide polymorphisms (SNPs), while neglecting larger-scale variation. Here, we present an approach that detects large deletions in soybean lines from standard GBS data by identifying genomic regions with lower-than-expected coverage. We tested this approach on 92 mutant soybean lines obtained through fast neutron mutagenesis of cultivar M92-220. Deletions identified following GBS were validated using comparative genomic hybridization (CGH) data obtained for 8 of the 92 lines. The GBS-based method detected 5 out of the 7 deletions observed by CGH and did not detect any deletion not found by CGH. Among the 92 individuals subjected to the GBS protocol, 111 high-confidence deletions (with mean and median size of 900 and 194 kb, respectively) were found. Preliminary results indicate that heterozygous deletions and duplications could also be found by this GBS approach, although with lower sensitivity than homozygous deletions. Our results suggest that GBS can be used to efficiently and cheaply detect deletions larger than 100 kb in size, with higher resolution in euchromatic than heterochromatic regions. This approach could be used to characterize a larger number of mutants from the same population so as to make full use of the potential of this genetic resource for functional genomics.

P0200: Methods: Sequencing

Genomic Introgression Accompanied by Recurrent Selection for Domestication Traits under the Single Origin of Cultivated Soybeans

Xutong Wang and Jianxin Ma, Purdue University, West Lafayette, IN
Soybean (*Glycine max* [L.] Merr.), which is one of the most economically important crop in the world, was domesticated from *Glycine soja* in China *ca.* 5,000 years ago as the wildly accepted hypothesis. The genetic admixture analyses showed that different degrees of introgression were detected in wild and domesticated accessions.
Despite the genetic diversity were dramatically lost (*ca.* 50%) due to breeding selection, selective sweep regions, which containing lots of domesticated candidate genes, were largely identified in whole genome of soybean. Although those genes may have played key roles in

soybean evolution and domestication history, only a few genes can be considered as domesticated genes, which also support a single origin hypothesis.
The single origin hypothesis, which consists in a statement that all domesticated soybeans derived from a single domestication event, was also shed light on using genome re-sequencing studies. Human-mediated introgression have played an important role in the genomic architecture of domestic animals and crops. For example, the existence of *GmHs1-1* in some landraces can be explained by genomic introgression between wild and domesticated soybean. Introgression might have been followed recurrent natural or domesticated selection in some selected genomic region.
Here, we detected the introgression signatures between wild and domesticated soybean in whole genome scales using whole genome re-sequencing data in 302 soybean accessions first time. Furthermore, we profiled the distribution feature of introgression fragments in whole genome and special selection regions. More specifically, we interrogated genotype variants associated with domesticated traits, to unravel the complexity of phenotype variations.

P0201: Methods: Sequencing

Next Generation Sequencing for Dissection of Drought Tolerance and Mining of Functional Markers in Tea

Rajni Parmar, Romit Seth, Pradeep Singh, Abhishek Bhandawat, Gopal Singh, Sanjay Kumar and Ram Kumar Sharma, Institute of Himalayan Bioresourse Technology, CSIR, Himachal Pradesh, India, Palampur, India

Drought is an environmental condition which negatively affects crop growth, yield and considered as significant cause of crop loss worldwide. Tea [*Camellia sinensis* (L.)] a perennial plant from Theaceae family is highly cross pollinated and heterozygous crop. It is consumed by billion of people worldwide due to its numerous health benefits; India is among the largest producer and exporter of tea in world. However, tea cultivation in India is significantly constraints due to impact of climate change leading to drought conditions, drastically affecting production by 14-33% and causes mortality up to 6-19%. Therefore, minimizing agricultural crisis caused by drought is major concern. High throughput transcriptome sequencing of diverse germplasm with contrasting response to drought gives valuable understanding of variable molecular response among/across different genotypes. Field evaluated potential four diverse drought tolerant and sensitive genotypes of tea were used for sequencing, *denovo* and reference-guided (tea genome) analysis was performed for identification of novel genes/transcripts. Approximately 93.9 million high quality reads were generated yielding 60,377 transcripts with average length of 1,074 bp. Functional annotation of 49,456 (80.3%) transcripts was achieved with multifarious public protein databases. KEGG and GO enrichment of differentially expressed genes among these contrasting genotypes identified 225 putative gene from five major categories viz., ABA-dependent and In-dependent pathway, Primary and Secondary metabolism, Ubiquitination, Transporters and Oxidative stress response may involved in providing tolerance. Additionally, 20,221 high quality non-synonymous SNPs and SSRs were ascertained from putative genes, are potential resource for genotyping, linkage map construction and QTL mapping in tea.

P0202: Methods: Sequencing

B-Carotene Differential Expression in Cassava: An Approach to Understanding High Accumulation in Roots

Tatiana Ovalle, CIAT, Cali, Colombia

Tatiana Ovalle, Manuel Ruiz, Sandra Salazar and Augusto Becerra López-Lavalle.

Malnutrition affects thousands of millions of people in the world, nearly two billion suffer from one or more micronutrient deficiencies. Mainly, because of critical crops to food security, contain low levels of micronutrients such as pro-vitamin A. Cassava (Manihot esculenta Crantz) is the staple food of more than 500 million people in the world.
Although continued efforts made by CIAT and Harvest Plus to improve the amounts of beta-carotene in the cassava roots is still low. Our goal is to identify differential expression in genotypes with different levels of β-carotene in roots and understand that proportion of genome is activated in high production of β-carotene. Current techniques, like RNA sequencing, allow elucidating the genes that contribute to features of interest. To do this, six contrasting genotypes for the β-carotene trait belong to a segregation family, and the two parents were evaluated phenotypically over three years, and RNA-sequencing was carried out.
Here, we reported the results of differential expression in contrasting genotypes to the β-carotene trait, and we show the regions up-regulated and down-regulated in cassava genome related to the accumulation of β-carotene in roots. Additionally, we compare expression levels in root and leaves tissue and we propose a hypothesis about the origin place and transport of β-carotene in cassava.
Keyword: RNA-seq, leave tissue, root tissue.

P0203: Methods: Sequencing

Differential Gene Expression during Flower and Fruit Development of the Blueberry Cv. O'neal

Rishi Aryal[1], Hamed Bostan[2], Ashley G. Yow[1], Elizabeth Tseng[3], Massimo Iorizzo[4] and Hamid Ashrafi[1], (1)North Carolina State University, Raleigh, NC, (2)Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, (3)Pacific Biosciences, Menlo Park, CA, (4)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC

Cultivated highbush blueberry (*Vaccinium corymbosum*) is one of the most recently domesticated crop plants. There is a high demand for improved blueberry cultivars due to its economic and nutritional value. Use of molecular breeding techniques can significantly expedite the conventional breeding process to fulfill the market demand. However, limited genomic resources are available to implement modern breeding techniques for blueberry. With the advent of next generation sequencing, it is now possible to sequence and explore polyploid and highly heterozygous genomes at a lower cost. One of the most important blueberry cultivars in North Carolina is cv. O'Neal, which is a tetraploid highbush blueberry species with unique characteristics such as good flavor, dry picking scar, and earliness. In this study, we compared the temporal change in transcriptome profile during flower development and fruit maturation of O'Neal. Total RNA was extracted from various stages of flower and fruit development as well as root and leaf tissues. Stranded mRNA libraries were sequenced by an Illumina HiSeq 4000. The same RNAs were also sequenced using PacBio Iso-Seq technology to study the isoform variation across the tissue types. A total of 1.8 billion Illumina reads were assembled into 219,842 contigs using CLC-Genomics. In total, 141,399 high quality sequences were obtained from 52 PacBio RSII SMRT cells. After collapsing similar isoforms using Cupcake ToFU package, 22,913 unique isoforms were retained. The

Illumina reads were mapped to transcriptome assembly and differential expression analysis was performed using the edgeR software package. The implication of using Illumina and PacBio Iso-seq data to study polyploid transcriptome variation will be discussed.

P0204: Methods: Sequencing

Generating High Quality Reference Genomes at the Vertebrate Genome Lab

Bettina Haase[1], Arang Rhie[2], Jacquelyn Mountcastle[1], Sadye Paez[1], Adam M. Phillippy[3], Erich Jarvis[4,5] and **Olivier Fedrigo**[1], (1)The Rockefeller University, New York, NY, (2)NHGRI/NIH, Bethesda, MD, (3)NHGRI, NIH, Frederick, MD, (4)Rockefeller University, New York, NY, (5)Howard Hughes Medical Institute, New York, NY

We have recently set up the Vertebrate Genome Laboratory (VGL) at the Rockefeller University as part of the Vertebrate Genome Project (VGP). The VGP is an international and multidisciplinary project of the Genome 10K (G10K) consortium in collaboration with Bird 10K, Bat1K and other consortiums. The mission of the VGP is to sequence at least one high quality near gapless, chromosomal level assembly of all ~66,000 extant vertebrate species. The primary outcome of the VGP is to create a digital Genome Ark to not only preserve genetic information but also provide tools for designing conservation strategies. Because sequencing technologies evolve rapidly, VGP members are continuously testing and evaluating new technologies and protocols. In Phase 1, the VGP is using a combination of 4 genome data types (Pacific Biosciences long reads, 10X Genomics linked reads, Bionano Genomics optical maps, and Hi-C chromosomal cross link reads) and up to two transcriptome types (illumina RNASeq or PacBio IsoSeq) in order to generate high-quality, phased, near chromosome-level, annotated, reference genome assemblies of species representing all vertebrate orders. The goal is to reach assembly quality greater than 1Mb contig N50, 10Mb scaffold N50, 90% of DNA assigned to chromosomes, QV40 base quality, and phased into haplotypes as much as possible. An assembly pipeline is being developed to handle these 4 datatypes. Once generated, the raw data and assemblies will be stored with DNANexus and Amazon Web Service. So far, using this pipeline, we have obtained assemblies with contig and scaffold N50 greater than 40Mb. Here, we present details, best practices, and preliminary results from the VGL data generation pipelines.

P0205: Methods: Sequencing

Development of Targeted GBS Panels for Breeding and Parentage Applications in Cattle and Swine

Angela Burrell[1], Michelle Swimley[1], Michelle Swimley[1], Roy C. Willis[1], Maarten de Groot[2], Ryan Ferretti[3] and Rick Conrad[1], (1)Thermo Fisher Scientific, Austin, TX, (2)VHLGenetics, Wageningen, Netherlands, (3)Neogen/GeneSeek, Lincoln, NE

Parentage testing and genomics-assisted breeding are critical aspects of successful herd management. Due to its highly accurate and reproducible results, targeted GBS is becoming an increasingly favored technology for SNP genotyping. With the utilization of next-generation sequencing, labs can test hundreds of samples across thousands of SNPs simultaneously in a simple high throughput workflow starting from either extracted nucleic acid or crude lysis samples.

We developed targeted sequencing panels for both cattle parentage, based on 200 SNP markers selected by the International Society of Animal Genetics (ISAG), and swine breeding using a 1500 SNP imputation panel. Utilizing the AgriSeq™ HTS Library Kit, a high-throughput targeted amplification and re-sequencing workflow, each panel's performance was tested on >96 diverse cattle and swine DNA samples. Libraries were sequenced on the Ion S5™ using an Ion 540™ chip with genotyping calling generated using the Torrent Variant Caller (TVC) plugin

The mean genotype call rate of markers across the samples was >98% for the cattle panel and >96% for the swine panel. Concordance across replicate library preparations and independent sequencing runs was >99.9% for both panels. Panel results were compared with results from a DNA array and the genotype call concordance was >99% with the AgriSeq workflows. The cattle panel was also used on field samples by a Netherland service lab to successfully determine the parentage relationships of 45 calves with 48 potential mother cows.

The data demonstrates the utility of the AgriSeq targeted GBS approach for cattle and swine SNP genotyping applications.

P0206: Methods: Sequencing

Identification of Genetic Variation Regulating Gene Expression in Dairy Cattle with RNA Sequence Data

Amanda J. Chamberlain[1], Ben J. Hayes[1,2], Ruidong Xiang[3], Christy J Vander Jagt[1], Coralie M. Reich[1], Iona M MacLeod[1], Claire P. Prowse-Wilkins[1], Brett A. Mason[1], Hans D. Daetwyler[1,4] and Michael E. Goddard[1,3], (1)Agriculture Victoria, Bundoora, Australia, (2)The University of Queensland, Brisbane, Australia, (3)The University of Melbourne, Parkville, Australia, (4)La Trobe University, Bundoora, Australia

There is increasing evidence to suggest many mutations affecting complex traits may be regulatory, that is they affect the expression of genes. The identification of regulatory variants could lead to increases in the accuracy of genomic breeding values. With the aim of identifying this type of variant, we performed three different analyses that tested variants for an association with changes in gene expression via three regulatory mechanisms: 1) By changing the total gene expression (expression QTL, eQTL), 2) changing the balance of the two parental alleles expressed (allele specific expression QTL, aseQTL), 3) changing the isoforms that are expressed (splice variant QTL, sQTL). We utilised RNA sequence data from milk and white blood cells collected from 141 lactating cows with imputed whole genome sequence data for all three analyses. Many variants were detected with significant eQTL, aseQTL and sQTL effects, with low false discovery rates. There was significant overlap in genes with significant eQTL, aseQTL and sQTL. eQTL with a large effect in white blood cells were likely to have a large effect, in the same direction, in milk cells as well. sQTL significant in both milk and white blood cells more often caused expression of the same isoform. There was a trend for the most significant variant to be < 100 kb from the transcription start site of the gene they were affecting for all three QTL types. The putative regulatory mutations affecting gene and isoform expression identified here are candidates for mutations affecting complex traits. Future work will refine these variants to those likely to be involved in regulating genes associated with traits important to the dairy industry and will test their impact on genomic prediction accuracy.

P0207: Methods: Sequencing

Decipher of the Spatiotemporal- and Sex-Related lncRNA Expression in Rhesus Macaque Brain

Xiangting Wang, University of Science & Technology of China, Hefei, China

Long noncoding RNAs (lncRNAs) represent a heterogeneous family of RNAs that are defined by length over 200 nucleotides and function as RNA molecules without protein encoding capacity. Although recent studies showed that certain individual lncRNAs play important roles in the central nervous system, many questions are still unresolved. In this work, we combined comprehensive analyses of RNA-seq and CAGE-seq (cap analysis of gene expression and sequencing) to decipher the dynamic changes of lncRNA expression in rhesus macaque brain. Four age groups from postnatal to aged periods, 8 brain regions and gender difference were applied. We identified 18 anatomically diverse lncRNA modules and 14 mRNA modules representing spatial, age, and sex specificity. Functional clustering of the interacted mRNAs and lncRNAs revealed that the lncRNAs were extensively involved in neuron related activities. Moreover, the CAGE-seq analysis found the usage of alternative promoter for both lncRNAs and mRNAs was dynamically changed in a spatiotemporal- and sex-related manner. All together, our findings offer an initial insight into spatial-, age- and sex-biased changes in lncRNA expression in macaque brain, and that the distinct classification of such changes might represent a previously unappreciated regulatory system which potentially contributes to brain development and aging.

P0208: Methods: Sequencing

High-Quality Assembly of European House Dust Mite Genome, Transcriptome and Proteome Reveal a Wide Range of Novel Allergens

Ming-Qiang WANG[1], Kevin Yi Yang[1], Carol Po-Ying Lau[1], Zeng Xi[1], Elaine Lai-Han Leung[2], Zhi-Gang Liu[3] and Stephen Kwok-Wing Tsui[1], (1)School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong, (2)State Key Laboratory of Quality Research in Chinese Medicine, Macau Institute for Applied Research in Medicine and Health, Macau University of Science and Technology, Macao, Macao, (3)State Key Laboratory of Respiratory Disease for Allergy at Shenzhen University, School of Medicine, Shenzhen University, Shenzhen, China

European house dust mite (HDM), scientifically known as *Dermatophagoides pteronyssinus*, is an especially important allergenic animal species. In this study, we aimed to construct a high-quality reference genome with the transcriptome and proteome methods to characterize novel allergen genes. A total of 62 Gb genomic data deposited with the BioProject number PRJNA388362 was generated from PacBio SEQUEL, Illumina HiSeq 2000 and Ion Torrent sequencing platforms. We performed *de novo* assembly, scaffolding, gap filling and polishing processes to obtain the 66.8 Mb assembly contains 1,390 contigs and 634 scaffolds, with scaffold and contig N50 being 194 kb and 80 kb, respectively. The assembly contains 8.9% repetitive sequences and 15,339 annotated protein-coding genes. The genome completeness as determined by BUSCOs analysis using 1,066 core genes from arthropoda_odb9 dataset was 89.1%. Our high-quality genome represents an 10-fold improvement in contig N50 compared to a previously published American HDM genome. In addition, gene structures of 21 canonical and 11 non-canonical allergens gene from *Der p 1* to *Der p 34* have been characterized with transcriptome data supported. We also identified 53 potential allergen protein sequences which were highly similar with allergens sequences of other species. Immunoblotting results showed that 50 HDM proteins that were bound by specific IgE in the sera of patients with HDM allergy. Those proteins were identified by MALDI-ToF MS and analyzed with the proteome predicted by the assembled genome. In summary, this study provides important genetic resources for further development of diagnostics and immunotherapeutic vaccines for HMD allergic patients.

P0209: Methods: Sequencing

Genomes of Pathogenic Fungi of *Bromus tectorum*

Nathan J Ricks, Brigham Young University, Provo, UT

Bromus tectorum (cheatgrass or downy brome) is an invasive, annual grass that now dominates large sectors of rangelands in Utah, Nevada, Idaho, Washington, Oregon and British Columbia. Native grass and shrub species have dwindled as a result of the introduction and expansion of this invader in the region. We have been studying the impact of several fungal pathogens on seed viability in the soil seed bank as well as vegetative growth of the plant. In conjunction with these studies we recently reported the genome sequence of *Pyrenophora semeniperda*, which has been shown to infect cheatgrass seeds in the soil. We have identified additional fungal pathogens associated with a cheatgrass stand failure phenomenon reported across the Great Basin region of the western United States. One of these fungi is an unidentified species of the *Rutstroemia* genus related to dollar spot (*Sclerotinia homeocarpa*) which infects above ground cheatgrass tissues, inducing bleaching and premature lodging of the plant. A second fungal pathogen associated with cheatgrass stand failure is a *Fusarium* species belonging to the *trincinctum* group, which, like *P. semeniperda*, infects cheatgrass seeds. To further advance our understanding of these important cheatgrass pathogens, we have now obtained genome sequence for three *Rutstroemia* sp. isolates and two *Fusarium* sp. isolates associated with cheatgrass stand failure. We have assembled and annotated these genomes with the goal of establishing their phylogenetic relationship to related species and furthering our understanding of genes associated with pathogenecity. The three *Rutstroemia* isolates yielded assemblies of approximately 41 Mb with 99 (N50 = 1.14 Mb), 79 (N50 = 970 kb) and 144 (N50 = 460 kb) contigs. The two *Fusarium* isolates had genome sizes of 90 Mb and 51 Mb with 2136 (N50 = 79 kb) and 5004 (N50 = 620 kb) contigs, respectively. The genomes were annotated using the Maker pipeline, yielding approximately 11,450 gene models for the three *Rutstroemia* isolates and an average of 17,000 genes for the two *Fusarium* genomes. Busco analysis demonstrated that each of the five assemblies contained more than 98% of the near-universal, single-copy orthologs associated with ascomycete fungal genomes.

P0210: Methods: Sequencing

A Multiple Amplicon Approach for Microbial Community Analysis on the Fluidigm Juno Platform

Mark Band, Chris L. Wright, Mark A. Mikel and Alvaro Hernandez, Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, Urbana, IL

With the introduction of Next Generation Sequencing, 16S ribosomal amplicon sequencing has been the most popular method for determining the distribution of bacterial strains among microbial communities. While today, shotgun sequencing is the gold standard, it remains a more expensive option and may not always be optimal, such as in samples with contaminating host DNA or low DNA concentrations. Here we present a multigene targeted approach that leverages NGS with microfluidic PCR, allowing simultaneous amplification of bacterial and

archaeal 16S, eukaryotic 18S, fungal ITS and functional-gene targets. The procedure results in barcoded, multi-amplicon libraries ready for sequencing on an Illumina MiSeq or HiSeq. Notably, we can also generate full-length fragments for use on Oxford Nanopore technology.
We have adapted the Fluidigm Juno System with LP48.48 Integrated Fluidic Circuits for automated amplification, barcoding, and linker addition in 2,304 individual amplification chambers. Following amplification, harvested products are quantified, run on a Fragment Analyzer to confirm sizes, pooled, and gel-size selected. Up to 1,536 individually-barcoded samples may be combined into a single pool for sequencing. The pool is quantitated by qPCR and sequenced on a MiSeq or HiSeq with 250nt or 300nt paired-reads based on the number of sample/primer combinations to target a minimum of 20k sequencing reads each.
Sequence data is parsed both by sample barcode and by primer set, and is then ready to be analyzed in QIIME, Mothur, IM-tornado, or other software packages.
This novel microbial profiling system is a fast, cost-effective, multi-targeted approach for whole microbial community analysis.

P0211: Methods: Transformation
iPLANTA: European RNAi Research Network
Bruno Mezzetti, Department of Agriculture - Marche Polytechnic University, Ancona, Italy and Jeremy Sweet, J T Environmental Consultants Ltd, Cambridge, Italy
Methods to exploit plant defence mechanisms or changing plant metabolism by RNA silencing show great potential. Interfering RNA can be used to improve plant composition while enhancing levels of beneficial nutrients, and to improve plant productivity by suppressing undesirable traits and switching resources to more beneficial quality and yield traits. The HORIZON2020 iPLANTA network (http://iplanta.univpm.it/), will define and coordinate the most important research tasks for the development of these novel transgenic strategies across 31 EU and nearby countries with inputs from cooperating researchers in Associated countries (N and S America, Australasia etc..). The project has the following main tasks: a) Evaluation of the efficacy of the RNA molecules for the induction of disease and pest resistance and metabolic changes. b) Examination of the specificity of the selected miRNAs and siRNAs and their impacts on both target and non-target/off-target systems. c) Developing specific risk assessment and risk management guidelines which relate to the data requirements and specific effects of the miRNAs and siRNAs on food, feed and the environment. d) Understanding the modes of transmission, uptake, systemic spread and degradation of dsRNAs, mi- and siRNAs. f) Determining the environmental and socio-economic impacts of plantRNAi technology and products and developing a public dissemination strategy.

P0212: Methods: Transformation
Plant Transformation and Genome-Editing Services at Donald Danforth Plant Science Center
Veena Veena, Preethi Vallal, Ying Rong and David Busby, Donald Danforth Plant Science Center, St. Louis, MO
Rising global population, economic growth in developing countries, and societal push for sourcing energy while preserving the environment and vital natural resources, are among greatest global challenges of the 21st century. Combinatorial approaches including efficient and sustainable crop production methods, improvement by breeding and genetic engineering will be required to meet these emerging challenges. Plant genetic engineering through the recent developments in targeted gene-editing technologies has highlighted the importance of plant transformation in multiple areas from functional genomics to crop improvement.
The Donald Danforth Plant Science Center is at the forefront in addressing these growing challenges through cutting-edge plant science research. **"The Plant Tissue Culture and Transformation Facility"** is an integral part of the Danforth Center's core research facilities located in St. Louis, Missouri. Over the years of its existence, facility has produced transgenic plants from tobacco, petunia, tomato, Arabidopsis, soybean, Indian mustard, maize, cassava, sweet potato, potato, and most recently green foxtail (*Setaria viridis*). The facility is enabling progress in plant research by development and optimization of transformation technologies in both model and crop plant systems, providing self-service training and access to state-of-the-art equipment and high quality space for researchers and companies use that are external to the Danforth Center.
The Plant Tissue Culture and Transformation Facility provides full-service research and consulting services to academic and commercial researchers external to Danforth Center, delivering transgenics and cell cultures utilizing our state-of-the-art equipment.

P0213: Methods: Transformation
***In Planta* Particle Bombardment: A Practical Way for Transforming Recalcitrant Crop Varieties**
Haruyasu Hamada[1], Qianyan Linghu[2], Yozo Nagira[1], Ryuji Miki[1], Naoaki Taoka[1] and **Ryozo Imai**[2], (1)KANEKA Corporation, Takasago, Japan, (2)NARO Institute of Agrobiological Sciences, Tsukuba, Japan
The currently favoured method for wheat (*Triticum aestivum* L.) transformation is inapplicable to many elite cultivars because it requires callus culture and regeneration. Here, we developed a simple, reproducible, *in planta* wheat transformation method using biolistic DNA delivery without callus culture or regeneration. Shoot apical meristems (SAMs) grown from dry imbibed seeds were exposed under a microscope and subjected to bombardment with different-sized gold particles coated with the *GFP* gene construct, introducing DNA into the L2 cell layer. Bombarded embryos were grown to mature, stably transformed T_0 plants and integration of the *GFP* gene into the genome was determined at the fifth leaf. Use of 0.6-µm particles and 1350-psi pressure resulted in dramatically increased maximum ratios of transient GFP expression in SAMs and transgene integration in the fifth leaf. The transgene was integrated into the germ cells of 62% of transformants, and was therefore inherited in the next generation. We successfully transformed the model wheat cultivar 'Fielder', as well as the recalcitrant Japanese elite cultivar 'Haruyokoi'. Our method could potentially be used to generate stable transgenic lines for a wide range of commercial wheat cultivars. We also discuss about the possiblity to apply the method to other crops.

P0214: Methods: Transformation
Transformation of Cowpea (*Vigna unguiculata* L.) Meristem Explants *via* Agrobacterium and Particle Bombardment
Edward Williams[1], Brian J. Martinell[1], Michael W. Petersen[1], Shawn Kaeppler[2], Heidi F. Kaeppler[3], Amy J. Miyamoto[1], Robert W Harnish[1], Vai Sa Nee Lor[4], Chantilly R. Schroedl[1], Seth Olson[1] and Abigail Ball[1], (1)Wisconsin Crop Innovation Center,

University of Wisconsin Madison, Middleton, WI, (2)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI, (3)Department of Agronomy, Wisconsin Crop Innovation Center, University of Wisconsin, Madison, WI, (4)Wisconsin Crop Innovation Center, University of Wisconsin Madison, Middleton, WI, Middleton, WI
Cowpea is consumed by over 200 million people in Africa daily, grown on over 10.5 million hectares worldwide, but susceptible to a variety of stresses that make it an attractive target for genetic modification. Cowpea transformation has been successfully reported using cotyledonary-node explants from mature seedlings using *bar* as a selectable marker, albeit at a low frequency. Direct transformation of meristem explants derived from seed offers efficiency gains and was first described in soybean. Here we describe an efficient method of transforming meristem explants of multiple cultivars of Cowpea directly isolated from seed.

P0215: Methods: Transformation

A Plant Regeneration Platform to Apply New Breeding Techniques for Improving Disease Resistance in Grapevine Rootstocks and Cultivars

Silvia Sabbadini[1], Luca Capriotti[2], Cecilia Limera[1], Oriano Navacchi[3] and **Bruno Mezzetti**[4], (1)Department of Agriculture, Marche Polytechnic University, Ancona, Italy, (2)Department of Agriculture, Marche Polythechinc University, Ancona, Italy, (3)Vitroplant Italia Srl, Cesena, Italy, (4)Department of Agriculture - Marche Polytechnic University, Ancona, Italy
Worldwide grapevine cultivation is based on the use of elite cultivars, in many cases strictly linked to local important wine brands. Most of *Vitis vinifera* cultivars have high susceptibility to fungal and viral diseases therefore, new breeding techniques (eg. Cisgenesis, RNAi and gene editing) offer the possibility to introduce new clones of the main cultivars with increased diseases resistance, in order to reduce environmental impact and improve quality in the intensive wine grape industry. This study is finalized to develop efficient *in vitro* regeneration and transformation protocols to extend the application of these technologies in wine grape cultivars and rootstocks. With this aim, *in vitro* regeneration protocols based on the production of meristematic bulks (Mezzetti et al, 2002) were optimized for different grapevine cultivars (Glera, Ciliegiolo, Thompson Seedless) and rootstocks (1103 Paulsen, 110 Richter and Kober 5BB). The meristematic bulks were then used as explants for *Agrobacterium* mediated genetic transformation protocols, by comparing the use of NPTII and e-GFP as marker genes. Results confirmed the efficiency of meristematic bulks as the regenerating tissue to produce new modified plants in all the above genotypes. The highest regeneration efficiency in some genotypes allowed the selection of stable modified lines with only the use of e-GFP marker gene. This protocol can be applied in the use of MYB marker gene for the production of cisgenic lines.

P0216: Methods: Transformation

Efficient Transformation of Williams82 and Elite Cultivars of Soybean

Michael W. Petersen[1], Brian J. Martinell[1], Edward Williams[1], Robert W Harnish[1], Amy J. Miyamoto[1], Jill Mahoy[1], Chantilly R. Schroedl[1], Seth Olson[1], Heidi F. Kaeppler[2] and Shawn Kaeppler[3], (1)Wisconsin Crop Innovation Center, University of Wisconsin Madison, Middleton, WI, (2)Department of Agronomy, Wisconsin Crop Innovation Center, University of Wisconsin, Madison, WI, (3)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI
Soybean transformation is a critically important technique for discovery research and advancing new traits of agronomic importance. Most methods of plant transformation and editing are reliant on older “transformation competent” germplasm, lengthy in duration, and are prone to tissue culture induced mutations. The method described here circumvents these issues by directly targeting the meristem of the developing seed. Here we describe a method of transforming Soybean explants derived from Williams82 and the Illinois elite genotypes 3025N and 3849N.

P0217: Methods: Transformation

Tissue Culture Induced Stochastic and Directed Changes in DNA Methylation in Maize.

Peter A. Crisp, Department of Plant and Microbial Biology, University of Minnesota, St Paul, MN, Zhaoxue Han, Northwest A&F University, Shaanxi, China, Scott Stelpflug, Monsanto Company, Huxley, IA, Shawn Kaeppler, Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI, Qing Li, National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Hubei, China and Nathan M. Springer, Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN
In some plant species tissue culture is widely used for clonal propagation of cultivars and in many other plant species tissue culture is required for successful transformation and regeneration of transgenic materials. While it is expected that plants derived from tissue culture will have no changes in genetic information, there are frequent examples of somaclonal variation, heritable phenotypic differences in plants recovered from tissue culture. Several locus-specific and genome-wide studies have provided clear evidence that DNA methylation patterns are perturbed by tissue culture. However, the mechanistic basis for the observed perturbations in DNA methylation remains largely unknown. In this study we utilized a sequence-capture bisulfite sequencing approach to profile context specific DNA methylation levels in 25 maize plants recovered from tissue culture. We find many regions with altered DNA methylation levels, particularly in the CG and CHG context. A subset of the changes in DNA methylation are consistently found in many independent events and appear to represent homozygous changes in DNA methylation that occur during tissue culture. Moreover, we document multiple examples of “Bad Karma like” epialleles with context-specific changes in DNA methylation. At these loci, heterochromatic DNA methylation is lost at transposable elements located within genes. This study may shed light on the mechanisms that drive epigenome changes during tissue-culture.

P0218: Methods: Transformation

Efficient Wheat Transformation Can be Performed on Cold-Conserved Immature Hybrid Embryos

Robin Michard[1], Marie-Claire Debote[1], Alain Loussert[1], Caroline Tassy[1], Pierre Barret[1], Giacomo Bastianelli[2], Alain Nicolas[3] and Pierre Sourdille[1], (1)INRA UMR 1095 GDEC, Clermont-Ferrand, France, (2)MEIOGENIX, Paris, France, (3)Institut Curie, Paris, France

Bread wheat is one of the three most cultivated crops in the world and a major economic challenge. However, since a few years, wheat production has reached a plateau. Despite the major role of conventional breeding programs in crop improvement, genetic engineering has become the fastest way to introduce new and well-characterized genes in plants leading researchers to elaborate each day new genetic transformation protocols. As wheat has become a new model plant for crop studying, especially in the word of genetic transformation, the need for an efficient protocol is without appeal. The present study aims to propose improvement of current immature embryo Biolistic® transformation protocols using cold conservation and hybrid immature embryos. We were able to show that using 4°C conservation for immature embryo storage do not affect regeneration and transformation efficiency. Moreover, using immature hybrid embryos can allow simultaneous transformation of two wheat genotypes, even if one of the genotypes is recalcitrant to genetic transformation. We think that those processes can be generalized to optimize wheat Biolistic® protocols.

P0219: Methods: Transformation

Development of Drought and Stress Tolerant Switchgrass (*Panicum virgatum* L) Cultivar *via Agrobacterium*-Mediated Transformation of a Bacterial Choline Oxidase a Gene (*CodA*)

Peiyu Zeng, State University of New York at Cobleskill, Cobleskill, NY and Lynda McMaster-Schuyler, State University of New York Cobleskill, Cobleskill, NY

Switchgrass (*Pabicum virgatum L.*) is a warm season (C4) perennial grass, native to North America which is capable of growing in poor soil conditions. It is an important forage crop and has received increased interest as a potential biofuel crop because of its high biomass. Water deficiency has been a major inhibitor of plant growth worldwide. Choline oxidase (COD) is an enzyme that catalyzes the production of glycinebetaine (GB). Build-up of GB in cells has been linked to drought and stress tolerance in bacteria and plants. The CodA gene codes for COD that originates from *Arthrobacter pascens.* The bacterial COD pathway is a single step pathway for GB production. It has been reported that the transgenic *Arabidopsis* and transgenic *Nicotiana tabacum* which are overexpressing the CodA gene show stress tolerance at all plant life cycle stages. In this project, we are overexpressing the CodA gene in switchgrass by *Agrobacterium*-mediated transformation. We have successfully produced transgenic plantlets and now are conducting molecular, physiological and progeny analysis.

P0220: Bioenergy

The Poplar Tree Microbiome: Implications of the Ecosystem Within

Sharon L. Doty[1], Andrew W. Sher[2], Pierre M. Joubert[1], Matthew Joseph[1], Andrea Firrincieli[3], Dehong Hu[4], Adam Deutschbauer[5], Zareen Khan[1] and Galya Orr[4], (1)University of Washington, Seattle, WA, (2)University of California, San Diego, La Jolla, CA, (3)University of Tuscia, Viterbo, Italy, (4)Pacific NW National Lab, Richland, WA, (5)Lawrence Berkeley Natl Lab, Berkeley, CA

With the relatively long life cycles of plants, symbiosis with microorganisms may allow plants to rapidly overcome environmental challenges. Endophytes are bacteria and fungi that live in intimate association within plants. The plant microbiota provide numerous benefits to the host plant including N-fixation, phytohormone production, reduced stress responses, antimicrobial production, tolerance to heat, salt, and drought, and pollutant degradation. Although some plant species are leguminous or actinorhizal, associating with rhizobia or Frankia, respectively, in root nodules, many pioneer plant species are non-nodulating and yet thrive in low-nutrient settings. For these plants, N-fixing (diazotrophic) endophytes and other closely associated microorganisms may be the source of this essential macronutrient. We study the diazotrophic endophytes of poplar (*Populus*) and willow (*Salix*), pioneer plant species able to colonize the rocky substrates deposited following riparian flooding. The $^{15}N_2$ incorporation assay was used to directly demonstrate N-fixation in cuttings of wild poplar. The microbiota varies considerably within the plants in number and in species with the intriguing possibility that only specific communities are active. Putative diazotrophic microorganisms were isolated from wild poplar plants, characterized, and sequenced. The presence of nitrogenase (*nif*) genes, in vitro assays for N-fixation activity, and quantitative *nifH* fli-FISH supported the hypothesis that the endophyte strains are diazotrophic. Random barcoded TnSeq is underway to elucidate the endophyte genes required for plant colonization and N-fixation. A consortium of the strains was added to hybrid poplar, increasing growth and N-fixation under greenhouse conditions. Not only do the microbes improve growth of this important bioenergy plant species, they also increase growth, health, and yields of an exceptionally broad range of plant species, including rice, tomato, pepper, strawberries, ryegrasses, and Douglas-fir. Inoculation of plants with endophytes improved water use efficiency and drought tolerance of the host plant. With the increased stress of climate change, the implications of plant-microbe symbioses for agriculture, forestry, and bioenergy production are profound.

P0221: Bioenergy

Comparative Analysis of Transcriptomes in Sugarcane Reveals Hormones' Role in Regulating Plant Growth.

Fan Zhu[1], Tyler Jones[2], Chifumi Nagai[2] and Ray Ming[1,3], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)Hawaii Agriculture Research Center, Kunia, HI, (3)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China

Sugarcane is a highly productive first generation biofuel feedstock, known for its remarkable efficiency in accumulating biomass. Hormones are important regulators for many biological processes in plants, especially in plant development and plant growth, which are crucial to plant biomass traits. To understand how hormones regulatory mechanisms contribute to sugarcane lignocellulose yield, we studied the transgressive biomass segregation in the F2 population derived from a cross between *Saccharum officinarum* 'LA Purple' and *Saccharum robustum* 'MOL5829'. Gene expression profiling was used to detect genes involved in three important hormone related pathways, auxin, ethylene and gibberellin, to find out how they are differently regulated between the extreme low biomass and the extreme high biomass group. We identified nineteen differentially expression genes in auxin, ethylene and gibberellin related signaling pathway, which could potentially regulate the biomass yield. Many genes families involved in auxin related signal pathways were differentially expressed, but only few genes were differentially expressed in the other two pathways. However, the genes involved in gibberellin and Ethylene pathway could play an important role in biomass accumulation. We also found that the differentially expressed genes PIF3 and EIN5 might involve in the crosstalk of ethylene signaling and gibberellin signaling pathway. These plant hormones related genes could potentially serve as candidate genes in genetic modification and breeding programs to develop high yielding energy cane.

P0222: Bioenergy

Potential Biofuel Relevant microRNAs Catalog for Sweet Sorghum Varieties through Deep Sequencing

Muhammad Y Khan[1], Ahmad Aziz[1], Muhammad Din[2] and Deependra Bhatta, (1)Tennessee State University, Nashville, TN, (2)University of Balochistan, Quetta, Pakistan

Sweet Sorghum (*Sorghum bicolor*) is an emerging, potential candidate crop for biofuel production due to its vast adaptability, low input cultivation, and high yield potentials. microRNAs (miRNAs) are the tiny molecular switches that play significant role in gene regulation of various cellular processes. The miRNAs profiles relevant to a developmental process may facilitate an approach to fine-tune that organism for desirable traits. This study aimed to profile and characterize the sorghum miRNAs obtained from leaf and stem and their targeted proteins by bioinformatics approaches and deep sequencing in four sweet sorghum varieties (two US varieties: Dale and Topper as well as two Pakistani varieties: Dasht Local and Achi Turi) at vegetative and reproductive stages. A total of 626 miRNAs with precursors were identified, where 401 were conserved while 225 were found to be novel miRNAs. Target prediction, Gene Ontology (GO) based functional characterization as well as Kyoto Encyclopedia of Genes and Genomes (KEGG) based functional enrichment were conducted on these miRNAs. The target proteins' determination as well as GO and KEGG functional analyses revealed that some miRNAs may have vital roles in biofuel production and biotic and abiotic stresses which can be helpful in devising better plant management tools.

P0223: Bioenergy

The Consortium for Advanced Sorghum Phenomics: Genome-Wide Association Mapping Using High-Density Drone-Based Phenotyping for Sorghum Bicolor Biomass Traits Under Drought Conditions

Jennifer Spindel[1], Jeffery Dahlberg[2], Matthew Colgan[3], Joy Hollingsworth[2], Julie Sievert[2], Hui Shen[4], Robert Hutmacher[5], Christer Jansson[6] and John P. Vogel[1], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)UC Kearney Agricultural Research and Extension Center, Parlier, CA, (3)Blue River Technology, Sunnyvale, CA, (4)Chromatin inc., Chicago, IL, (5)West Side Research & Extension Center, Five Points, CA, (6)Pacific Northwest National Laboratory: Environmental Molecular Sciences Laboratory, Richland, WA

By the end of 2050, production of all staple cereal grains including sorghum will need to double to meet the needs of a global population of ~9 billion people. At the same time, agriculture is imperiled world-wide by climate change, and there is a dire need for renewable fuels. The Consortium for Advanced Sorghum Phenomics (CASP) is a large, inter-disciplinary DOE-funded project with the goal of accelerating the breeding of biomass sorghum adapted to drought stress conditions. Over the summer of 2016, 622 diverse, public, *Sorghum bicolor* conversion lines and 26 proprietary sorghum cultivars provided by Chromatin Inc. were grown in two locations in central California under well-watered, pre-flowering drought stress, and post-flowering drought stress conditions. All plots were phenotyped weekly for biomass, leaf area index, and plant height by Blue River Technologies using a drone-based phenotyping platform. Using a custom genome-wide association study pipeline, we identified over 200 reliable and conserved associations for biomass traits, which speaks directly to the power of drone-based phenomics for gene discovery.

P0224: Bioenergy

Metabolomic and Transcriptomic Analyses of Upland and Lowland Switchgrass Rhizomes Transitioning to Winter Dormancy

Nathan A. Palmer[1], Erin D. Scully[2], Gautam Sarath[1], Christian Tobias[3], Serge Edme[1], Rob Mitchell[1], Scott Sattler[1] and Soundararajan Madhavan[4], (1)USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE, (2)USDA-ARS Center for Grain and Animal Health Research, Manhattan, KS, (3)USDA-ARS, Western Regional Research Center, Albany, CA, (4)University of Nebraska-Lincoln, Lincoln, NE

Perenniality of switchgrass (*Panicum virgatum*) resides in the rhizomes and crown regions of the plant. New tillers initiated from these perenniating structures yields the new season's biomass. Although some of the physiological aspects of transition to dormancy have been studied in switchgrass, there is essentially a complete lack of data on the molecular aspects of these key processes. For this study, rhizomes were collected from field-planted cv Summer (upland ecotype) and cv Kanlow (lowland ecotype) switchgrass plants and analyzed for a range of metabolites, plant hormones, and by RNA-Seq. Rhizomes from cv Summer were collected over two consecutive years and used develop a system model for plants adapted to the Central Great Plains transitioning to winter dormancy. Kanlow rhizomes were collected concurrently with the first year of Summer samples and contrasted with the Summer-derived system model. Transcriptomic and metabolic similarities and differences in rhizomes harvested from these two ecotypes over the course of the growing season, with an emphasis on changes occurring during dormancy, are presented.

P0225: Bioenergy

Comparison of Genomic Selection Methods for Predicting Miscanthus sinensis Yield

Maria Stefanie Dwiyanti[1], Lindsay V. Clark[2], Kossonou Anzoua[3], Joe E. Brummer[4], Katarzyna Glowacka[5], Megan Hall[6], Kweon Heo[7], Xiaoli Jin[8], Alexander E. Lipka[9], Junhua Peng[10], Ji Hye Yoo[7], Chang Yeon Yu[7], Hua Zhao[11], Stephen Long[9], Toshihiko Yamada[3] and Erik J. Sacks[2], (1)Hokkaido University, Sapporo, Japan, (2)Dept of Crop Sciences, University of Illinois, Urbana, IL, (3)Hokkaido University, Sapporo, Hokkaido, Japan, (4)Colorado State University, Fort Collins, CO, (5)Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland, (6)Bio Architecture Lab, Berkeley, CA, (7)Kangwon National University, Chuncheon, Korea, Republic of (South), (8)Zheijiang University, Hangzhou, China, (9)University of Illinois, Urbana, IL, (10)Life Sci & Tech Center of China Seed, Wuhan, Hubei, China, (11)Huazhong Agricultural University, Wuhan, China

Increasing yield of bioenergy crop *Miscanthus* is one of its top breeding priorities. Genomic selection (GS) is a method whole-genome data to predict the breeding value of offspring, and it is particularly useful to predict complex traits like yield. While GS itself can predict yield, adding information from specific traits that are known to have large effect to yield will increase the GS accuracy. In this study, we compared three

methods: a standard GS model (RR-BLUP) where all markers were treated as non-fixed effect; GS+GWAS where significant markers from GWAS are used as fixed effect; and GS+selection index, using information from traits that have high genetic correlations with yield.

P0226: Bioenergy

Comparison of Transcriptomic Changes in *Miscanthus sinensis* under Salt Stress

Minjung Yook, Soo-Hyun Lim, Gyoungju Nah and **Do-Soon Kim**, Seoul National University, Seoul, South Korea

Miscanthus is generally known for its high abiotic stress tolerance, but has not been much studied. To cultivate *Miscanthus* in marginal lands including salt-stressed area, we need to understand the response mechanism to salt stress and how salt tolerance can be improved. Therefore, this study was conducted to explore global changes in the leaf transcriptome of two *M. sinensis* accessions with contrasting salt tolerance and to uncover which genes and regulatory mechanism related to salt tolerance of *M. sinensis.* A total of 363 DEGs were obtained from the pair-wise comparison of the six cDNA libraries. For salt-sensitive accession (M119), 143 genes were up-regulated during the first 24 hours after salt treatment, and keep the state of regulation in the subsequent 24 hours. For salt-tolerant accession (M131), 178 genes were sharply up-regulated during the first 24 hours and turned to the normal state in the subsequent 24 hours. Sixty seven DEGs (48 annotated) were exclusively up-regulated in salt-tolerant accession and most of them were associated with signal transduction pathway to abiotic stresses in plants including transcription factors and receptor-like kinases. Thus, our results suggest that gene expression regulations of the upstream genes in the salt-tolerance cascades brought about diverse adaptability for salt stress in *M. sinensis*. The result of this study could be utilized for studying and improving salt stress tolerance in Miscanthus.

P0227: Bioenergy

Global Transcriptome Analysis of *Salicornia Bigelovii*, an Oilseed Halophyte, Seedling Under Wastewater.

Houda Chelaifa, Masdar Institute, A Part of Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates, Youssef Idaghdour, Biology Department / Division of Science and Mathematics/ New York University Abu Dhabi (NYU AD), Abu Dhabi, United Arab Emirates, Massar Dieng, NYU Abu DHABI, Abu DHABI, United Arab Emirates, Ayesha Al Marzooqi, Biomedical Engineering/ Masdar Institute/ A Part of Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates and Hector H. Hernandez, Department of Biomedical Engineering/ Masdar Institute/ A Part of Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Salicornia bigelovii (Amaranthaceae) is a succulent annual C4 halophyte native to coastal salt marshes of Eastern and Southern America and Mexico. *S. bigelovii* has distinctive characteristics that make it a potential candidate for use as a model salt tolerant crop. In this context, *S. bigelovii* is used as an oilseed and biomass crop for treatment of saline aquaculture effluent of an integrated Seawater Energy and Agricultural System (SEAS). Despite, *S. bigelovii's* unique high salt stress tolerance, there is a paucity of work towards elucidating its transcriptome dynamics.

To characterize the global transcriptome of *S. bigelovii*, RNA sequencing was performed on four tissues (shoots-roots-flowers and seeds) during plant development to gain an understanding of transcriptome dynamics in response to aquaculture effluent irrigation. A total of 191 million reads were obtained using the Illumina HiSeq platform. De novo assembly using Trinity generated 643,752 high-quality transcripts with minimum length size of 200 bp. Further filtration with 5FPKM in 2 replicates generating 66,943 transcripts. The data show a high number of tissue specific differentially expressed genes, with a distinct gene expression profile present in the seeds transcriptome.

Analysis by qTR-PCR was performed using 7 genes implicated in ion homeostasis and cation transporters in order to validate the expression profiles observed in our Illumina sequencing data. The analysis confirmed the significant differential expression pattern of all 7 genes, suggesting that osmotic activity increases in roots and shoots and decreases in seeds and flowers to protect halophyte plant from salt damage to ensure plant propagation.

P0228: Bioenergy

Evaluation of Marker-Assisted Selection and Genomic Selection in Polyploid Backcross Populations

Alexander E. Lipka[1], Lindsay V. Clark[2], Xiping Yang[3], Jianping Wang[3], D.K. Lee[4], Claire Liu[5], Keiji Motomura[6], Sushma Sood[7], Vanessa S. Gordon[8], Toshihiko Yamada[9] and Erik J. Sacks[2], (1)University of Illinois, Urbana, IL, (2)Dept of Crop Sciences, University of Illinois, Urbana, IL, (3)Agronomy Department, University of Florida, Gainesville, FL, (4)University of Illinois Urbana-Champaign, Urbana, IL, (5)Syngenta, Ames, IA, (6)University of the Ryukyus, Nakagami District, Japan, (7)USDA-ARS, CANAL POINT, FL, (8)USDA-ARS Sugarcane Field Station, Canal Point, FL, (9)Hokkaido University, Sapporo, Hokkaido, Japan

One of the most promising approaches for introducing novel genetic diversity into polyploid plant breeding material is to introgress favorable alleles for a quantitative trait from a donor population to a set of elite taxa. The use of marker-assisted selection (MAS) in recurrent backcross populations to predict which offspring have optimal phenotypes could expedite such endeavors because breeders would not have to obtain phenotypic data from multi-year replicated field trials. Crucial to the success of MAS is the use of accurate statistical models relating genotype to phenotype. Depending on the genetic architecture underlying a quantitative trait of interest, one could make phenotypic predictions based on either a subset of markers tagging peak quantitative trait nucleotides (QTN) or a genome-wide marker set included in a genomic prediction model. The purpose of this work is to simulate traits with contrasting genetic architectures in biparental crosses from multiple polyploid plant species and then to contrast the ability of markers tagging QTN to predict trait values to that using genome-wide marker sets. In addition to using k-fold cross validation to assess prediction accuracy, we will also evaluate the phenotypic performance of subsequent backcross populations derived from individuals with optimal predicted phenotypes from either one or both of these marker sets. We expect these results to provide guidance on when it would be most appropriate to focus resources towards obtaining whole-genome marker sets versus sequencing only regions in the vicinity surrounding peak QTN.

P0229: Other Category

The Vertebrate Genome Project: A Digital Genome Library for the 66,000 Vertebrates

Sadye Paez[1], Jacquelyn Mountcastle[1], Bettina Haase[1], Olivier Fedrigo[1] and Erich Jarvis[1,2], (1)The Rockefeller University, New York, NY, (2)Howard Hughes Medical Institute, New York, NY

The Vertebrate Genome Project (VGP) is an international and multidisciplinary project led by the Genome 10K consortium with the Bird 10K, Bat1K, and other consortiums, with VGP partners in over 50 institutions in nearly all continents. The VGP aims to create a digital open-access genome library of at least one high-quality near gapless, annotated chromosomal level assembly of all ~66,000 extant vertebrate species. This library will be used to identify species most genetically at risk for extinction, to preserve genetic information, and to transform biology through conservation strategies as a model to preserve all life.

A major industry challenge for saving species has been finding cost-effective and sufficient technology to generate high-quality genomes. We worked with industry partners to develop unprecedented high-resolution genome sequencing methods at significantly lower costs than current, less robust technologies. Instead of so-called “swiss-cheese” genomes with many holes, we can now generate “whole cheese” genomes, assembled into maternal and paternal chromosomes, at a fraction of the cost. Our novel genome assembly solution has been validated on 7 species representing vertebrate diversity: a hummingbird, songbird, endangered parrot, sea bass, goat, endangered gorilla, and human. Despite possessing different genome structures, the technique successfully created chromosome assemblies with unprecedented resolution.

Earth is in the midst of its 6th mass extinction event, the worst since the die-off of dinosaurs 66 million years ago. However, this one is caused by human activities, including habitat destruction, pollution, climate change, and more, and the planet is at risk for losing at least 1 in 8 endangered vertebrate species (7,978 total) to extinction on the IUCN Red List. The VGP presents a viable and necessary solution towards addressing this event and becoming better stewards of the planet and our species’ survival.

P0230: Other Category

Developing a US STEM Workforce that is Globally Competitive

Marcela Yepes[1], Stephanie Fuchs[2], Matthieu Fuchs[2], Aleksey Zimin[3,4], Keithanne Mockaitis[5], James A Yorke[6] and Herb Aldwinckle[7], (1)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (2)Cornell University/ Biological Engineering, Ithaca, NY, (3)Institute for Physical Science and Technology, University of Maryland, College Park, MD, (4)Johns Hopkins University, Department of Computer Science, Baltimore, MD, (5)National Center for Genome Analysis Support, Pervasive Technology Institute, Bloomington, IN, (6)University of Maryland, College Park, MD, (7)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY

As part of our NSF funded project, we organize yearly a STEM career exploration workshop. The participants are rising Undergraduates and High School minority students as well as biology/science instructors and teachers. The program is designed to explore the intersection of science, technology, engineering and math (STEM) as it pertains to applications in biological research including big data analysis, gene editing, genomics, high through put sequencing, among others. Our NSF funded STEM exploration workshop is designed for undergraduate and high school students who may be interested to explore applications of STEM paths to medicine and biological research with emphasis in genomics and new cutting-edge science. Discussions and workshop exercises allow students to gain insight into applications of STEM while they interact with faculty and guest speakers. Students have the opportunity to gain a broad overview of STEM applications while they learn about current biological and medical cutting research topics that are at the intersection of medicine, biology, and engineering career paths that will equip students to lead in increasingly complex, interconnected, and diverse fields. The primary workshop objective is to give participants a greater understanding of interdisciplinary fields and motivate students to follow STEM paths to produce future leaders prepared to influence their communities and the world in positive ways. Exposure to STEM is critical for high school students and for inspiring future scientists in the US. It’s that spark, the discovery of what science and technology have to offer them in the future.

This abstract will be also presented as an oral presentation in the Coffee Genomics Workshop by coauthors M. Yepes, S. Fuchs, M. Fuchs, NSF Award Award#1444893

P0231: Other Category

P3: Predictive Plant Phenomics Graduate Training Program

Lyle A. Sisson, Ted Heindel, Carolyn J. Lawrence-Dill, Patrick S. Schnable and Julie Dickerson, Iowa State University, Ames, IA

Modern engineering and data analysis techniques make it feasible to develop methods to predict plant growth and productivity based on genome and environment information, however broader skillsets will be needed to unlock this potential, so student training methods must adapt. This poster describes the structure and activities of a National Science Foundation Graduate Research Traineeship (NRT) award focused on Predictive Plant Phenomics (P3). Our program aims to increase agronomic output as highlighted by the National Plant Genome Initiative’s current five-year plan [NST, 2014]. Ph.D. training production levels and types are not always a good fit for addressing complex technical and societal problems such as these. To train these scientists, the P3 NRT is using the T-training model proposed by the American Society of Plant Biology (ASPB) and described in “Unleashing a Decade of Innovation in Plant Science: A Vision for 2015-2025”. This approach requires that students get broad exposure to multiple disciplines, work with industry, and develop effective communication and collaboration skills - all without increasing the time to graduation. This poster describes how we are working towards meeting these challenges. Initial results show that the P3 students have more contact with faculty across departments than single discipline graduate students and are open to learning about new areas. However, we are still grappling with some issues like finding the best mechanism for balancing student skills through leveling activities such as boot camps and introductory course training early on in the program. To learn more about the P3 NRT, visit: https://www.predictivephenomicsinplants.iastate.edu.

P0232: Other Category

Integrating and Displaying Plant Gene Expression in Expression Atlas

Laura Huerta, Suhaib Mohammed, Nuno Fonseca, Maria Keays, Amy Tang, Elisabet Barrera, Wojtek Bazant, Anja Fullgrabe, Nancy George, Haider Iqbal, Alfonso Munoz-Pomer, Alvis Brazma, Robert Petryszak and Irene Papatheodorou, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Expression Atlas (https://www.ebi.ac.uk/gxa) is a database and web-service at EMBL-EBI that selects, curates, re-analyses and displays gene expression data in a baseline context, e.g. to find genes expressed in different tissues in potato, and in a differential context, e.g. to find up-regulated genes in response to stripe rust and powdery mildew in wheat. Plant experiments from ArrayExpress, GEO and SRA/ENA/DDBJ are selected for curation and analysis. Data curation involves enriching sample annotation with additional metadata, annotating metadata with Experimental Factor Ontology (EFO) terms and deciding comparisons for differential expression analysis based on associated publications and correspondence with the original researchers. Data analysis is performed using open source tools for microarray data and our standardized pipeline iRAP (https://github.com/nunofonseca/irap) for RNA-seq data.
Currently, we provide gene expression analysis results for more than 700 plant experiments across 20 different plant species. Expression Atlas can be searched by gene, gene set and biological condition queries. The use of EFO annotations allows efficient search via ontology-driven query expansion and facilitates data integration across multiple experiments. We offer downstream analysis and visualization such as gene co-expression, biological variation among replicates, transcript quantification, visualization of gene expression in Gramene genome browser and enrichment of Gene Ontology terms and Reactome pathways. Finally, we have developed an automatic pipeline that discovers new plant RNA-seq data at ENA for 45 different species, performs quality control, alignment to the genome reference in Ensembl plants and quantification of gene and exon expression. The analysis results are available via our RNASeq-er API (https://www.ebi.ac.uk/fg/rnaseq/api/).

P0233: Other Category
The International Alliance for Phytobiomes Research
Kellye Eversole, International Phytobiomes Alliance, Bethesda, MD
Crop productivity in sustainable agricultural systems must improve considerably to meet the demands of a global human population expected to reach 9.8 billion by 2050. Traditional cropping systems will have to produce record levels of food, feed and fiber while facing a steadily changing climate and increased biotic and abiotic stressors.
To optimize productivity, sustainability, and profitability on farms, grasslands, and forests, scientists must embrace and focus on the complexity within phytobiomes. A phytobiome is a plant ("phyto") in a distinct geographical unit ("biome") – a field, grassland, or forest. It includes the plant itself, all micro- and macro-organisms living in, on, or around the plant – such as microbes, animals and other plants – and the environment, including soil, air, water, weather, and climate. Until now, agricultural sciences have relied mostly on research within individual disciplines and reductionist approaches for crop improvement and production methods and practices. This must change.
By focusing on the phytobiome, we will be able to elucidate, quantify, model, predict, act, manipulate, prevent, and ultimately prescribe the cropping systems, methods, and management practices most suited for sustainable production on a particular farm, grassland, or forest.
The Phytobiomes Alliance is an international, nonprofit consortium of academic institutions, large and small companies, and governmental agencies. The goal of the Alliance is to develop a foundation of knowledge on phytobiomes and to translate that knowledge into applications for farmers, ranchers, and foresters. An overview of the goals and priorities of the Phytobiomes Alliance will be presented.

P0234: Other Category
Engineering Speciation: A New Tool for Transgene Biocontainment and Pest Biocontrol
Maciej B. Maselko[1], Siba R. Das[1], Stephen C. Heinsch[1], Jeremy M. Chacon[2], Aidan J. Peterson[3], James E. Parker[4], Michael B. O'Connor[3], Przemyslaw G. Bajer[5], William R. Harcombe[2] and Michael J. Smanski[1], (1)Department of Biochemistry, Molecular Biology, and Biophysics and BioTechnology Institute, University of Minnesota, St. Paul, MN, (2)Department of Ecology, Evolution, and Behavior and BioTechnology Institute, University of Minnesota, St. Paul, MN, (3)Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, (4)Computer Science and Engineering, University of Minnesota, Minneapolis, MN, (5)Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN
We present data that demonstrate the feasibility of Synthetic Incompatibility (SI); a method of engineering species-like barriers in sexually reproducing organisms. SI utilizes genome editing to introduce a silent mutation in a conserved region of a promoter followed by the expression of programmable transcriptional activator targeted to the non-edited sequence. Hybridization between an edited organism expressing the programmable transcriptional activator and an organism with a non-edited promoter results in lethal overexpression caused by transcriptional activation of the non-edited promoter. Animal applications of SI include genetic biocontrol of pest species, replacing disease vector populations with engineered non-vector organisms, preventing gene flow between genetically engineered fish or livestock and their non modified counterparts as well as the genetic biocontainment of experimental systems such as gene-drives. Agricultural applications include preventing transgene flow from engineered crops which may enable the use of herbicide resistant traits in plants with closely related weeds and large scale production of crops engineered to make high-value compounds. Results from proof of principle experiments in *Saccharomyces cerevisiae* and *Drosophila melanogaster* will be presented.

P0235: Other Category
Introducing Genomics through Gene Annotation: Broadening Access to Course-Based Undergraduate Research Experiences
Sarah C R Elgin, Washington University in St Louis, St. Louis, MO and the Genomics Education Partnership, G-OnRamp, and the Genomics Education Alliance
The Genomics Education Partnership (GEP; http://gep.wustl.edu) introduces undergraduates to genomics/bioinformatics by engaging them in a research project. Currently students are working to improve the genome sequence quality and annotate the small, heterochromatic "dot" chromosome (F element) and a comparable euchromatic portion of the D element in a group of *Drosophila* species. Our goal is to use phylogenetic footprinting to identify organizational features and regulatory elements that promote gene expression in the heterochromatic (F element) environment. Assessments show that GEP students from very diverse schools learn about genes and genomes, and gain an appreciation of the research process. To broaden the scope of scientific projects undertaken, GEP has partnered with Galaxy (http://galaxyproject.org) to produce G-OnRamp (http://gonramp.org -see for workshops) — a Galaxy server enabling biologists with little/no

informatics expertise to generate a UCSC Assembly Hub or a JBrowse/Apollo genome browser for a newly sequenced genome, with sequence similarity, *ab initio* gene predictions, genomic repeats, and RNA-seq evidence tracks. GEP faculty have begun to use the browsers produced by G-OnRamp to study the evolution of biochemical pathways, initially triglyceride production in parasitoid wasps. Given the growing importance of genomics in the life sciences, members of several programs that provide course-based undergraduate research experiences (GEP, GCAT-SEEK, ComGen, Ciliate Genomics Consortium, Genome Solver, CSHL DNA Learning Center, etc.) have proposed the formation of a Genomics Education Alliance https://figshare.com/articles/A_Genomics_Education_Alliance/5197228). We invite other groups interested in sharing resources and participating in outreach to join us. Supported by NSF IUSE 1431407 (GEP) and NIH R25GM119157 (G-OnRamp).

P0236: Other Category

Training Next Generation Plant Breeders: Uganda's Model Regional Training Program at Makerere University

Patrick O Ongom[1], Paul T Gibson[1], Richard Edema[2], Isaac Onziga Dramadri[2], Bruno Awio[1] and Boris Mahule Eyse Awio[1], (1)Makerere University Kampala, Kampala, Uganda, (2)Makerere University, Kampala, Uganda

Africa has significant shortage of topnotch plant breeding professionals. The knowledge of plant breeding remains the nucleus for developing and deploying crop cultivars suited to the needs of human. To meet future food demand in Africa, we need to train more expert plant breeders. Makerere University Regional center for Crop Improvement (MaRCCI) has developed a unique approach to graduate plant-breeding training. The initiative was initially supported by AGRA (Alliance for a Green revolution in Africa) and recently by the World Bank. The program is designed to deliver theoretical and practical skills, enabled by two model breeding programs: cowpea and sorghum, and a biotechnology unit that the center support. Training emphasis is on applied plant breeding and molecular skills, and other soft skills oriented toward managing plant breeding projects. Our curriculum is enhanced by usage of an E-learning platform (Plant Breeding E-Learning in Africa) developed by Iowa State University. Student research projects are designed in collaboration with: National agricultural research Organization, CIAT and IITA whose scientists mentor the students. Cooperation with national/international research organizations, and regional and US universities (Iowa State, Cornell, North Carolina State, Purdue University, University of California Riverside) have strengthened the program. The skills gained allow our graduates to independently run breeding programs, meeting smallholder and crop value-chain needs in the African context. Since 2008, the center has been able to train several African graduates at Masters (117) and PhD (58) levels from over 20 sub-Saharan countries, who are now running breeding programs in the region.

P0237: Other Category

"On-Demand" Trait Enablement in Crops using Florian™ Switch Technology

Arianne Tremblay, AgBio Intrexon Corp, Davis, CA and Amanda Edwards, Rio Stamler, Trang Dang, Patrick Canlas, Jyoti Rout, Shiv Tiwari, Sekhar Boddupalli

Over the past 3 decades, agricultural biotechnology has revolutionized food and feed productivity with commercialization of crop protection and quality traits. Several of the crop protection traits are based on over-expression of transgenic pesticidal proteins. Rapid adaptation of these traits along with year-round crop production in several geographies has led to resistance development that is imminent across the globe. In addition, constitutive over production of the pesticidal proteins could negatively impact plant vigor and be yield limiting in some cases. In view of this, "on-demand" expression of transgenic traits, combined with the increasing adoption of precision agricultural tools could be particularly advantageous for future crop pipelines. In order to explore the potential of this concept, we have developed and tested Florian™ switch technology, a system which enables inducible control of gene expression using cost-effective and scalable chemistries for broad agricultural applications. Specific examples demonstrating flowering control, crop protection, and quality traits in two different plant systems will be presented. Potential commercial applications to further the Florian™ switch technology include: a) increasing biomass and feed quality in forage crops by prolonging the vegetative stage, b) enabling "on-demand" seed production with improved efficiency and effectiveness, c) synchronizing flowering in high value fruit and produce to aid in harvest timing (e.g. strawberries, pineapples), and d) bringing flexibility in crop protection against biotic and abiotic stresses based on disease pressures or climate conditions.

P0238: Other Category

AgBioData: Identifying and Meeting Common Goals in the Management of Agricultural Biological Data

Jacqueline D. Campbell[1], Ethalinda Cannon[1], Sook Jung[2], Dorrie Main[2], Monica Poelchau[3], **Ramona Walls**[4] and Lisa C Harper[5], (1)Iowa State University, Ames, IA, (2)Washington State University, Pullman, WA, (3)USDA/Agricultural Resarch Service/National Agricultural Library, Beltsville, MD, (4)University of Arizona, Tucson, AZ, (5)USDA-ARS, Albany, CA

The future of agricultural research depends on data. The sheer volume of agricultural biological data being produced today makes good data management essential. Governmental agencies, publishers and science funders now require data management plans for publicly funded experiments. Furthermore, the value of data increases exponentially when they are properly stored and shared, so that they can be easily utilized in future analyses. AgBioData is a consortium of agricultural biological databases which strive to identify common goals relating to data set acquisition, display, user retrieval and manipulation; data, software and hardware standards, and database best practices. The objectives of AgBioData include: 1) coordinate with external groups such as the Genomics Standards Consortium (GSC), CyVerse, and scientific journals to develop standards for efficient data flow among data generators and databases, 2) encourage communication and sharing between databases, to identify common problems and collaborate on solving them, 3) encourage sharing between databases to avoid duplication of work and support small research groups with data management. Ultimately, we aim to solve common issues to both leverage our work and to create database products that are more findable, accessible, interoperable, and reusable. Towards this end, we have written an extensive white paper covering biocuration, communication, interoperable GGB (genetics, genomics, and breeding) platforms, data sharing, metadata and persistence, and ontologies, the contents of which will be described in this poster.

P0239: Other Category

Key Laboratory of Resource Biology and Biotechnology in Western China

Guifang Zhao, Northwest University, Xi'an, China

Quercus is one of the most important genera for considering its economic and ecological values, with approximately 500 species worldwide. Recently, the oak phylogenetic backbone has been well established based on pollen characteristics and nuclear markers, and six major intrageneric groups (the Old World Clade: *Cyclobalanopsis*, *Cerris* and *Ilex*; the New World Clade: *Lobatae*, *Protobalanus*, and *Quercus*) have been identified. However, there still existed ambiguous relationships within the groups of *Querucs*. Estimating the phylogenetic relationships among *Quercus* has been particularly difficult when considering the species diversity and extensively hybridization. Currently, plastid genome-scale data has achieved great progress in inferring difficult phylogenies. China is a second center of *Quercus* diversity but with less intensively evolutionary studied. Here, with 50 species representing all four groups (*Cyclobalanopsis*, *Cerris*, *Ilex* and *Quercus*) that are distributed in China, we explored the phylogenetic relationships among these oaks based on the plastid genome sequences. Our results showed that high resolution was achieved within most clades based on the exons of protein-coding genes. The phylogenetic tree revealed that group *Quercus* formed a sister clade to the only representative (*Quercus rubra*) of group *Lobatae* with strong support. The remaining oaks formed another branch. Notably, groups *Cyclobalanopsis* and *Cerris* were found nested in group *Ilex*, rendering the group *Ilex* non-monophyletic. The scenario was also observed in the previous studies that inferring from few plastid markers. Incomplete lineage sorting or introgression of oaks may give rise to the phenomenon. Additionally, the result may also be related to the slow evolutionary rates of plastid genomes. Overall, our plastid genome data provided strong support for most relationships despite with some uncertain nodes, and it would be necessary to explore the unambiguous *Quercus* phylogeny by combining plastid data with nuclear data in the future study.
Keywords: *Quercus*, China, Plastid phylogenomics, Non-monophyletic

P0240: Other Category

Unleashing Genetic Diversity by Increasing Meiotic Recombination : An *in silico* Benchmark

Elise Tourrette, **Matthieu Falque** and Olivier C. Martin, GQE– Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France

Genetic diversity is the fodder of genetic progress during selection processes, whether natural or artificial. Recombination during meiosis generates genetic diversity via the formation of crossovers which thus shuffle allelic combinations. Crossovers are heterogeneously distributed along chromosomes (hot and cold regions). This is important because for example in maize, 30% of the genes are in very cold regions and they arise with significant polymorphism. Selection can be impeded if advantageous alleles arise in such cold regions as they are thus difficult to extract or separate from nearby disadvantageous alleles.

We have been studying the influence of recombination rate, recombination landscape and other parameters (selection intensity, population size, genetic architecture such as the distribution and effects of QTLs) on the behavior of genetic gains and diversity when populations are subject to recurrent selection in breeding programs.

Our approach is based on simulations using a quantitative genetics framework. We focus on plant breeding in *Arabidopsis*, cacao, barley and maize, investigating different ways of modifying recombination, based on what is known to be possible experimentally today: modification of the genome-wide recombination rate (HyperRec technology), mainly acting on cold pericentromeric regions, and targetting hot spots to specific parts of the genome.

To extract information about the behavior of the population along the breeding program, different observables are measured, including genetic gain, genetic diversity, linkage disequilibrium and also the coupling/repulsion status between loci. With our simulations, we identify how the different breeding choices influence the genetic gains achieved across successive generations for different breeding choices and scenarios for modifying recombination.

P0241: Other Category

Prevalence, Mechanisms and Importance of Duplicate Gene Divergence in Exon–Intron Structure

Xuehao Fu, Hongyan Shan, Peipei Wang, **Guixia Xu** and Hongzhi Kong, Institute of Botany, Chinese Academy of Sciences, Beijing, China

Gene duplication plays key roles in organism and genome evolution. Understanding how duplicate genes evolve and diverge through time is critical for elucidating the mechanisms underlying the origins of new characters and new organisms. Previous studies have shown that, at least in plants, some duplicate genes have diverged in the exon-intron organization, suggestive of structural divergence. However, because the species and gene pairs sampled were very limited, it is still unclear whether this phenomenon is widespread and, if yes, how prevalent it is. In this study, by conducting a genome-wide study on closely related duplicate genes from four representative species of plants (*Arabidopsis thaliana*), animals (*Drosophila melanogaster*), fungi (*Saccharomyces cerevisiae*), and protists (*Paramecium tetraurelia*), we found that structural divergence occurred prevalently in every examined species but with different proportions, ranging from 70.9% in *P. tetraurelia* to 91.9% in *A. thaliana*. Three mechanisms, including exon/intron gain/loss, exonization/pseudoexonization, and intra-exonic insertion/deletion, are detected to be responsible for structural divergence. Similar to non-synonymous substitutions, the probability of duplicate genes to diverge in structure increases with evolutionary time. We used Pearson correlation coefficient (PCC) and Euclidean distance (ED) to evaluate the difference of expression patterns between a pair of genes, and observed that PCC is lower and ED is higher in diverged than in undiverged duplicate genes, suggesting that structural divergence may be coupled with expression divergence. Using d_N/d_S values as a measurement of functional constraint, we found that duplicate genes with structural changes have higher d_N/d_S values, indicative of weaker functional constraint on these genes. This is in concordance with their lower expression levels. Further function enrichment analysis revealed that these genes are involved in unconserved biological processes, including biological regulation, signal transduction, and response to stimuli. Our findings show that the modes of structural divergence of duplicate genes are generally consistent in different eukaryotic species, implying that structural divergence is an important contributor to the evolution of duplicate genes.

P0242: Other Category

Single Cell Transcriptomics for Characterization of Complex Cellular Systems

Alvaro Martinez Barrio[1], Paul Ryvkin[2], Zachary Bent[3], Stephane Boutet[2], Sofia Kyriazopoulou-Panagiotopoulou[2], Josephine Lee[3], Patrick Marks[2], Samuel Marrs[3], Elliot Meer[4], Jeff Mellen[2], Luz Montesclaros[2], Daniel Riordan[5], Joe Shuga[2], Jessica Terry[2],

Kevin Wu[2], Grace Zheng[2], Tarjei Mikkelsen[6] and Deanna M. Church[2], (1)10X Genomics, Pleasanton, CA, (2)10x Genomics Inc., Pleasanton, CA, (3)10x genomics, Inc., (4)10x genomics, Inc, (5)10x Genomics, Inc, (6)10x Genomics, Inc.
The advent of high-throughput, droplet-based systems for assaying transcriptomes at single cell resolution has revolutionized our approach to studying complex biological systems. We have developed a fully-integrated, droplet-based approach, the ChromiumTM single cell system, that enables unbiased expression profiling of single cells. High efficiency cell capture coupled with a low doublet rate (<1% per 1000 cells) facilitates the profiling of precious and rare cell populations. To support this system, we provide an open source analysis pipeline, Cell RangerTM, and an interactive data visualization tool, LoupeTM Cell Browser.
We demonstrate the power of this system to characterize the subpopulations of the murine embryonic brain. Starting with ~1.3 million brain cells from cortex, hippocampus and ventricular zones of 2 E18 mice we generated over 100 single cell libraries, sequencing each cell to ~18k raw reads. Major neuronal and non-neuronal cell types from different brain layers were detected, including diverse, yet rare interneurons. Recent system upgrades include improved analysis pipelines and expanded chemistry for capturing both unbiased gene expression and VDJ receptor pairing in T and B cells. We used the upgraded system to profile the tumor microenvironment and investigate innate immune response in three different tumor types.
Chromium single cell library preparation protocols have been demonstrated for starting material from multiple sample types, including human, murine and moss (https://support.10xgenomics.com/single-cell-gene-expression/sample-prep); protocols with additional non-model systems are under development. We envision that our integrated single cell platform will enable applications in a wide variety of systems and organisms, and accelerate the characterization of diverse biological systems.

P0243: Aquaculture

The National Center for Biotechnology (NCBI) Genome Annotation Resources for Aquaculture Species

Nuala A. O'Leary, Francoise Thibaud-Nissen, Terence D. Murphy and Eukaryotic Genome Annotation Team, National Center for Biotechnology Information (NCBI), Bethesda, MD
NCBI's Eukaryotic genome annotation pipeline (ncbi.nlm.nih.gov/genome/annotation_euk/) incorporates genomic, transcript, and protein sequence records, including RNA-seq data available in SRA, to provide comprehensive annotations of public genome assemblies submitted to NCBI's Assembly resource (ncbi.nlm.nih.gov/assembly). To date, this pipeline has been used to annotate more than 420 eukaryotic genomes across diverse array of taxa. Among these annotated genomes are several economically important aquaculture species including catfish, salmon, tilapia, scallop and oyster. The annotations provided by this pipeline are available in various NCBI resources, including Reference Sequence (RefSeq) sequence databases, Gene, BLAST databases, FTP and in NCBI's Genome Data Viewer. All genome annotations produced by this pipeline are in scope for manual curation by the RefSeq curation group. Curators correct sequence or feature annotation errors that are identified by quality assurance tests, generate additional splice variants, and add feature annotation and data attributes. These curated RefSeq transcript, proteins and genomic regions, designated by NM_NP_, NR_, or NG_ accession prefixes, serve as reagents to NCBI's Eukaryotic genome annotation pipeline, thereby contributing iteratively to improvements in our genome annotation products. This presentation will describe some of the computational and manual curation procedures used in NCBI's genome annotation process and provide guidance on how the resources can be accessed and utilized by the aquaculture research community.

P0244: Aquaculture

High Quality Genome Assemblies of Emerging Evo-Devo Models: The Zebrafish and Its Relatives

Kerstin Howe[1], William Chow[1], Shane McCarthy[2], Zemin Ning[1] and Richard Durbin[2], (1)The Wellcome Trust Sanger Institute, Cambridge, United Kingdom, (2)University of Cambridge, Cambridge, United Kingdom
The zebrafish and its close relatives, the Danioninae, are an emerging model system for the study of evolution and development. Within the Vertebrate Genomes Project (VGP) aiming at sequencing and assembling all vertebrates to chromosome level, we have already started to address a selection of the Danioninae.
The project was launched in early 2017, and has so far seen 17 *Danio rerio* strains/populations, 7 additional *Danio* species and 2 other close relatives added to the sequencing pipeline. Particular attention is being paid to generating a haplotype-resolved high-quality assembly of the SAT strain, a cross between double-haploid parents from Tuebingen and AB and therefore an ideal model for method development.
We are applying a variety of sequencing and assembly technologies, including chromosome-level scaffolding with RACA, and quality-check and improve the resulting assemblies using the Genome Assembly Evaluation Browser gEVAL (geval.sanger.ac.uk) which was extended to feature a bespoke analysis pipeline for VGP assemblies. The Danioninae genomes are not without challenges, and so far, many routine approaches have failed due to high heterozygosity and repeat content. We are describing what is being done to address these issues, and how we reconcile and evaluate data from a variety of sequencing and mapping resources and assembling/scaffolding software.

P0245: Aquaculture

Chemokine *ccl33* Is a Key Regulator of Teleost Fish Barbel Development

Tao Zhou, Ning Li, Yulin Jin, Qifan Zeng, Wendy Prabowo, Yang Liu, Changxu Tian, Lisui Bao, Shikai Liu, Zihao Yuan, Qiang Fu, Sen Gao, Dongya Gao, Rex Dunham and Zhanjiang Liu, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL
Barbel is integumentary sense organ that can be found in fishes, reptiles and amphibians. It is one of the most characteristic of catfish (Siluriformes), a diverse vertebrate order with approximately 4,000 species. However, barbel-less catfish does exist such as bottlenose catfish (*Ageneiosus Marmoratus*), which serves as an ideal model to determine the genomic basis for barbel development and its regeneration. In this project, we produced the whole genome reference genome sequence of the bottlenose catfish, and transcriptome of the barbel tissue. We also identified a set of genes differentially expressed during barbel regeneration. Through comparative genome and transcriptome subtraction analysis, we were able to pin down the candidate gene that is present only in barbeled fish but not in barbel-less fish, and also differentially expressed during barbel regeneration. Here we report a chemokine gene, *ccl33*, as a key regulator of barbel development. It is present in the genomes of fish with barbels but absent from genomes of fish without barbels. It has multiple copies in channel catfish (*Ictalurus punctatus*), of

which three copies, *ccl33.4*, *ccl33.6*, and *ccl33.7* were differentially expressed during barbel regeneration, in a fashion concordant with the timing of barbel regeneration. Knockout of *ccl33* genes in zebrafish (*Danio rerio*) resulted in various phenotypes including complete loss of barbels, reduced barbel sizes, curly barbels, and weak performance. Similarly, knockout of *ccl33* in channel catfish caused partially or completely loss of barbels, and body deformity. These results suggest that *ccl33* is a key regulator of barbel development, and may be important for normal development of sensory immunological functions. In addition, this study also demonstrated the power of comparative subtraction of genomes and transcriptomes for the determination of genomic basis for genes involved in specific traits, especially those for morphological traits.

P0246: Aquaculture

Pedigree Assessment for Polygamous Fish Using SNPs

Ajith Premachandra Horane Karayalage, University of the Sunshine Coast, Maroochydore, Australia, Nguyen Hong Nguyen, University of the Sunshine Coast, Maroochydore DC, Australia and Wayne Knibb, University of the Sunshine Coast, Sippy Downs, Australia

Accurate pedigree assignment is often a fundamental requirement for applied breeding programs. However, pedigree assignments can be challenging for mass spawning, polygamous species due to the large number of half sib families arising from mass spawning, so highly informative markers are required in order to achieve a high level of accuracy of pedigree assignment. In the present study, the parentage and kinship predictions were obtained for 622 individuals, using single nucleotide polymorphism (SNP) and the results were compared with those obtained using eight previously described DNA microsatellite markers in yellowtail kingfish (*Seriola lalandi*), which is a popular food-fish with high aquaculture importance in worldwide. A panel of 207 SNPs (mean H_o = 0.497) was able to assign the parental pairs for 96% of the offspring similar to the microsatellites (mean H_o = 717). For the full-sib assignment (without parental genotypes), over 97% of the full-sib pairs predicted using 207 SNPs were similar to that predicted from microsatellites. Increasing the number of SNPs, increased the number of full-sib pair combinations but the level of similarity of full-sib pairs between SNPs and microsatellites did not changed. Whereas, reducing the number of SNPs used in analysis drastically reduced both number of full-sib pair combinations and the similarity of predicted full-sib pair combinations between SNPs and microsatellites. These results suggest that novel SNPs can be successfully used for pedigree assignments in socially polygamous yellowtail kingfish and power of 207 SNPs appears to be more or less similar to the eight microsatellites panel.

P0247: Aquaculture

Identification of Sexually Differentially Methylated Regions in Channel Catfish Provides Evidence of Epigenetic Control of Its Sex Determination

Yujia Yang[1], Yang Liu[1], Lisui Bao[1], Changxu Tian[1], Zihao Yuan[1], Dongya Gao[1], Rex Dunham[1] and Zhanjiang Liu[2], (1)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (2)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY

Channel catfish is a dominant aquaculture species in the United States. It is known to have a XY sex determination system, but environmental factors such as temperature can affect sex phenotypes. Although genotypes are determined at the time of fertilization, its sex phenotypes are uncommitted before 19 days post fertilization (dpf). In this study, we conducted methylation mapping analysis in genotypic female and male channel catfish before the commitment of sex phenotypes at 9, 12, and 16 dpf. A total of 2,683 sexually differentially methylated CpG sites were identified. Interestingly, 51.28%, 34.01%, and 42.95% of differentially methylated CpG sites were located on sex chromosome (chromosome 4) at 9, 12, and 16 dpf, respectively. The sex control region had the highest density of methylated sites, spanning a physical distance of approximately 5 Mb (Chr4:15Mb-20Mb). This region was significantly more hyper-methylated in females than in males, suggesting epigenetic regulation of sex control. A total of 1,271 genes were annotated nearby differentially methylated CpG sites, many of which had functions associated with sex determination, sex chromosome evolution, gonadogenesis, and gonad differentiation. Detailed analysis of methylation patterns within a set of genes related to sex determination such as *idh2*, *sema4b*, *chd2*, and *rasgrf1* was conducted. Preliminary results suggested that intragenic regions and promoters were drastically differentially methylated between females and males. Along with analysis of sex determination genes, this work provides insights into the mechanisms of sex determination, and provides evidence that epigenetic control is involved in the sex determination in channel catfish.

P0248: Aquaculture

Identification, Phylogeny and Expression Analysis of FoxO Transcription Factors in Channel Catfish

Lei Gao[1,2], Zihao Yuan[3], Dongya Gao[3], Rex Dunham[3] and Zhanjiang Liu[4], (1)Auburn University, Auburn, AL, (2)Liaoning Ocean and Fisheries Science Research Institute, Dalian, Liaoning, China, (3)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (4)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY

FoxO (Forkhead box O) transcription factors are well known for their roles in immune responses in model organisms, including their regulation of antioxidant defense, cell apoptosis and cell cycles. However, very little is known about the FoxO transcription factors in aquaculture species. In this study, seven FoxO genes, including FoxO1a, FoxO1b, FoxO3a, FoxO3b, FoxO4, FoxO6a and FoxO6b, were identified and characterized from the channel catfish transcriptome and genome. Through phylogenetic and syntenic analysis, FoxO1, FoxO3 and FoxO6 were observed to be duplicated in zebrafish and catfish compared with other species, indicating a teleost specific gene expansion. In addition, the expression patterns of FoxO genes in response to bacterial (*Edwardsiella ictaluri* and *Flavobacterium columnare*) invasions were determined using meta-analysis with RNA-Seq datasets. Most of the FoxO genes were up-regulated after bacterial challenges, of which expression of FoxO6b was drastically induced after infection of *E. ictaluri*. Although expression of this gene was upregulated after infection of *F. columnare* as well, but was much less dramatic, suggesting that the induced expression was pathogen-specific. Taken together, the FoxO transcription factors appeared to be involved in immune responses of channel catfish after bacterial infection.

P0249: Aquaculture

Identification of QTL and Candidate Genes Associated with the Resistance for Enteric Septicemia of Catfish Using Genome-Wide Association Analysis in Channel Catfish

Huitong Shi[1], Tao Zhou[1], Xiaozhu Wang[1], Yujia Yang[1], Shikai Liu[1], Lisui Bao[1], Ning Li[1], Zihao Yuan[1], Yulin Jin[1], Suxu Tan[1], Dongya Gao[1], Rex Dunham[1] and Zhanjiang Liu[2], (1)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (2)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY

Disease resistance is one of the most important traits for aquaculture industry. Enteric septicemia of catfish (ESC), caused by the bacterial pathogen *Edwardsiella ictaluri*, causes enormous economic losses for the domestic catfish industry every year. However, molecular mechanisms of disease resistance to ESC are still not clear. In this study, three significant quantitative trait loci (QTL), with two of them located on LG1 and one on LG26, and three suggestive QTL located on LG1, LG3, and LG21, respectively, were identified to be associated with ESC resistance by genome-wide association study (GWAS) using the 690K catfish SNP arrays. With a well assembled reference genome sequence, genes around the involved QTL regions can be readily identified. On these genes, 37 genes had known functions in immunity, making them potential candidate genes for ESC resistance in channel catfish. Notably, *nlrc3* and *nlrp12* were also reported for ESC resistance in hybrid catfish, suggesting this QTL was operating within channel catfish populations and within interspecific hybrid populations. Many of the genes with functions in immunity were involved in Class I MHC pathway of the mediated antigen processing and presentation, indicating that this pathway was significantly associated with ESC resistance in channel catfish. This study validated one QTL previously identified using the fourth generation of the backcross progeny populations and F2 generation of backcross progenies, and identified five additional QTL among channel catfish families. Taken together, it appears that there are only a few major QTL for ESC disease resistance, making marker-assisted selection an effective approach for genetic improvements of ESC resistance.

P0250: Aquaculture

JAK and STAT Members in Channel Catfish: Identification, Phylogenetic Analysis and Expression Profiling after Bacterial Infection

Yulin Jin, Tao Zhou, Ning Li, Shikai Liu, Xiaoyan Xu, **Jian Luo**, Ying Pan, Suxu Tan, Huitong Shi, Yujia Yang, Zihao Yuan, Wenwen Wang, Rex Dunham and Zhanjiang Liu, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

The Janus kinase/signal transducers and activators of transcription (JAK/STAT) signaling pathway is one of the main pleiotropic cascades used to transmit information from extracellular receptors to the nucleus, which resulting in DNA transcription and expression of genes involved in immunity, proliferation, differentiation, migration, apoptosis, and cell survival. Specifically, members of JAK family and STAT family have been extensively studied in different mammalian species because their important roles in orchestrating mammalian innate and adaptive immune responses. However, they have not been systematically studied among teleost fish species. In this study, five JAK family members and eight STAT family members were identified and characterized from channel catfish genome by conducting comprehensive bioinformatic studies. Phylogenetic analysis verified their identities and showed the channel catfish JAK and STAT genes had closest relationship to zebrafish. Further syntenic analysis confirmed the phylogenetic analysis and suggested that JAK/STAT genes were well conserved among vertebrates. Compared to mammals, more members of the JAK and STAT family were identified in channel catfish genome. Expressions of JAK and STAT family members were detected in healthy catfish tissues with various abundances, reflecting their basic and important biological functions in normal cells. Moreover, for the first time, we determined the expression profiles of the JAK and STAT members after E.ictaluri infection. All of these members showed a general trend of up-regulation in gill, liver, intestine following bacteria challenge. Notably, the significant upregulation of STAT1b gene in catfish liver, gill and intestine after *E. ictaluri* infection support the notion that high STAT1 expression are involved in defense against pathogens. Collectively, the increased expression of JAK and STAT members in tested tissue suggested their crucial function in defending against pathogen invasions. Further studies are certainly worthwhile to elucidate the details of JAK and STAT members involved in regulation of immune response in fish species.

P0251: Aquaculture

Genomic Evaluation for Harvest Weight and Residual Carcass Weight in Channel Catfish Using Single-Step Genomic BLUP

André L. S. Garcia[1], Brian Bosworth[2], Geoff Waldbieser[3], Shogo Tsuruta[1], Ignacy Misztal[1] and Daniela A.L. Lourenco[1], (1)University of Georgia, Athens, GA, (2)USDA-ARS Warmwater Aquaculture Research Unit, Stoneville, MS, (3)USDA - Agricultural Research Service, Stoneville, MS

Catfish production is the largest aquaculture segment in the US. Since 2006 selection has been based on traditional BLUP evaluations and with the recent availability of genomic information the objectives of this study were: to investigate the feasibility of using genomic selection in US catfish and to identify major SNP associated with harvest weight and residual carcass weight. Phenotypes were available for harvest weight (n=27,160) and residual carcass weight (n=6020), and the number of fish in the pedigree was 36,365. After quality control, genotypes on 54,837 SNPs were available for 2911 fish. Genomic and pedigree predictions were calculated in a 5-fold cross validation approach, using single-trait models. Single-step genomic BLUP (ssGBLUP) was the method of choice for genomic predictions. Ability to predict breeding values was calculated as the correlation between adjusted phenotypes based on complete data and EBV or genomic EBV (GEBV). Inflation was assessed as the regression coefficient (b1) of adjusted phenotype on (G)EBV. The GEBV were back solved to SNP effects and the percentage of variance explained by each SNP was calculated as SNP effect squared. Predictive ability for both traits increased 8 points and bias was reduced when genomic information was used. The proportion of variance explained by windows of 20 SNP was at maximum 2.2% for harvest weight and 3.3% for residual carcass weight. Both traits appear to be polygenic with no major SNP. Using genomic information is beneficial in catfish selection because of higher predictive abilities and it also allows to identify superior individuals within families.

P0252: Aquaculture

GWAS Analysis of QTL for Resistance Against *Edwardsiella ictaluri* in F2 Interspecific Hybrid Catfish

Suxu Tan, Tao Zhou, Wenwen Wang, Yulin Jin, Xiaozhu Wang, Dongya Gao, Rex Dunham and Zhanjiang Liu, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL
Infectious diseases pose significant threats to the catfish industry. Enteric septicemia of catfish (ESC) caused by *Edwardsiella ictaluri* is the most devastating disease for catfish aquaculture, causing tens of millions of dollars of economic losses annually. Channel catfish and blue catfish exhibit variation in resistance against ESC, and as such the interspecific hybrid backcross progenies provide an ideal system for the analysis of the resistance QTL. In this study, we conducted GWAS analysis to locate genomic regions associated with ESC resistance by genotyping phenotype extremes of F2 backcross families with the catfish 690K SNP arrays. Two genomic regions on LG1 and LG23 were determined to be associated with ESC resistance as revealed by a mixed linear model and family-based association test. A number of genes within these QTL regions have known functions in immunity, making them potential candidates as disease resistance genes. For instance, seven genes on LG1 (*nck1*, *agtr1*, *trpc1*, *abi1*, *apbb1ip*, *actr3b*, and *vav3*) and three genes on LG23 (*mrc1l*, *prkcq*, and *gata3*) were involved in immune-related functions. These genes mainly function in signaling pathways of phagocytosis and T-cell activation, suggesting their roles in disease resistance. This study demonstrated the power of GWAS analysis for the identification of QTL in the hybrid system. We previously reported one QTL in LG1 using fourth generation of backcross families, which was validated here in the F2 backcross families, suggesting that this QTL is operating in various populations of a broad genetic background, making it useful for application in marker-assisted selection.

P0253: Aquaculture

The Annotation and Transcriptional Landscape of Repetitive Elements after Biotic and Abiotic Stress in Catfish

Zihao Yuan[1], Tao Zhou[1], Changxu Tian[1], Lisui Bao[1], Shikai Liu[1], Huitong Shi[1], Yujia Yang[1], Dongya Gao[1], Rex Dunham[1], Geoff Waldbieser[2] and Zhanjiang Liu[3], (1)Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL, (2)USDA - Agricultural Research Service, Stoneville, MS, (3)Syracuse University; Department of Biology, College of Art and Sciences, Syracuse, NY
Channel catfish (*Ictalurus punctatus*) is the most important aquaculture species in the United States and has been used as a model species for comparative immunology, physiology, and toxicology. As such, its reference genome was generated and annotated with protein coding genes. However, the repetitive elements in the catfish genome are poorly understood. In this study, over 417 Mb of repetitive elements were identified and characterized in the channel catfish genome. Among them, the DNA/TcMar-Tc1 transposons are most abundant, making up ~20% of the total repetitive elements, followed by microsatellites (14%), LINE/L2 (4.3%), Xba elements (3.6%) and LTR/Nagro (3.1%). The novel repetitive elements were classified into 216 novel categories. A number of catfish-specific repetitive elements were identified including the previously reported Xba elements. The expression profile of the transposable elements in catfish after bacterial infections of *Edwardsiella ictaluri*, *Flavobacterium columnare* and abiotic stress (hypoxia and heat stress) in different tissues were determined by meta-analysis of RNA-Seq datasets. Overall, more transposable elements were upregulated with infection of *Flavobacterium columnare* than with *Edwardsiella ictaluri*; heat stress caused drastic downregulation of transposable elements, significantly more so than those under hypoxic conditions. The mechanisms of the observed differential expression were unknown, but were believed to be associated with a specific set of genes that were regulated under the stresses.

P0254: Aquaculture

Development of a 50K Transcribed Gene SNP-Chip Identifies Major QTL for Growth in Rainbow Trout

Rafet Al-Tobasei[1], **Ali R. Ali**[2], Yniv Palti[3], Timothy D. Leeds[3] and Mohamed (Moh) Salem[2], (1)University of Alabama at Birmingham, Birmingham, AL, (2)Middle Tennessee State University, Murfreesboro, TN, (3)USDA-ARS-NCCCWA, Kearneysville, WV
Coding/functional SNPs change the biological function of a gene, therefor, may lead to identification of causative alleles within QTLs and development of genetic markers with large-effects on phenotypes. Two bioinformatics pipelines, GATK and SAMtools, were used to identify ~21K coding/functional SNPs with allelic-imbalances associated with important aquaculture production traits including BW, muscle yield, muscle fat content, shear force, and whiteness in addition to resistance/susceptibility to bacterial cold-water disease (BCWD). SNPs were identified from pooled RNA-Seq data collected from ~620 fish, representing 98 families from a growth- and 54 families from a BCWD-selected genetic lines with divergence phenotypes. In addition, ~29K SNPs without allelic-imbalances were strategically added to build a 50K Affymetrix SNP-chip. SNPs selected included 2 SNPs per gene from 14K genes and ~5K non-synonymous SNPs. The SNP-chip was used to genotype 1728 fish. The average SNP calling-rate for samples passing QC (1641 fish) was ≥ 98.5%. GWAS analysis on 783 fish (representing 200 families from 2 generations) X 40K polymorphic markers (passing QC) identified genomic loci on chromosome 9, 13 and 4 explaining 6-16% of the genetic variance of body-weight-gain. PLINK analysis on the same set of data identified 1783 SNPs significantly associated with body-weight-gain (P-value <4.99E-08). Majority of the SNPs identified by GWAS and PLINK had allelic imbalances with BW in the original SNP data set used to build the SNP-chip indicating high success rate of the bioinformatics pipelines in calling informative SNPs with allelic-imbalances from pooled samples and utility of the SNP-chip in GWAS studies in rainbow trout.

P0255: Aquaculture

Evaluation of the Utility of the New Rainbow Trout Genome Assembly for Analyzing RNA-Seq Data from Stress Response Experiments

Guangtu Gao[1], Gregory M. Weber[1], Sixin Liu[1], Yniv Palti[1] and Caird E. Rexroad III[2], (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)USDA, Beltsville, MD
The newly released rainbow trout genome assembly in NCBI RefSeq has greatly expanded our abilities for analyzing rainbow trout sequencing data. In this poster, we evaluate the utility of this genome assembly for analyzing RNA sequencing (RNA-seq) data of rainbow trout responses to various stressors, including high temperature, low temperature, re-use water quality, salinity, and confinement. A spliced-alignment software, STAR was used to align the RNA-seq reads to the reference genome. Aligned reads were counted using htseq-count, and DESeq2 was used in comparing the treatments with the controls. A total of 90% of the RNA-seq reads were mapped to the reference genome with ~75% and ~15% of the reads uniquely and multiply mapped to the genome, respectively. This is a major improvement from previous RNA-seq

studies in rainbow trout that mapped the RNA-seq reads to a reference transcriptome and would typically have only ~ 55% of the reads aligned to the reference. The majority of the reads were mapped to genome locations annotated by RefSeq gene models, but there was also a substantial amount of the reads that were mapped to genomic regions that are currently not annotated. Comparing with the control samples, differentially expressed genes (DEG) were found in the responses to all stressors but the confinement stress experiment. Functional annotations of the DEGs was conducted to identify molecular and biological pathways that may be involved in the rainbow trout stress response.

P0256: Aquaculture

GWAS for Detecting QTL Associated with Columnaris Disease in Two Rainbow Trout Breeding Populations

Rafael M.O. Silva[1,2], Jason P. Evenhuis[2], Roger L. Vallejo[2], Guangtu Gao[2], Kyle E. Martin[3], Ignacy Misztal[1], Timothy D. Leeds[2], Daniela A.L. Lourenco[1] and Yniv Palti[2], (1)University of Georgia, Athens, GA, (2)USDA-ARS-NCCCWA, Kearneysville, WV, (3)Troutlodge, Inc., Sumner, WA

The purpose of this study was to prospect genomic regions that explain large portion of the additive genetic variance for resistance against Columnaris disease (CD) in rainbow trout. Two important aquaculture populations were investigated. The National Center for Cool and Cold Water Aquaculture (NCCCWA) odd-year line, which was previously selected for bacterial cold water disease resistance; and the Troutlodge, Inc., May odd-year (TLUM) nucleus breeding population. The number of fish in the pedigree was 54,350 and 36,265, respectively; in which 8,453 and 3,986 fish had phenotypes recorded for CD resistance, respectively. Fish that survived to 21 days post immersion challenge were recorded as resistant. Genotypes for 57k SNPs (Affymetrix Axiom®) were available for 1,185 and 1,137 fish from NCCCWA and TLUM, respectively. The SNP effects and variances were estimated using the weighted single-step genomic BLUP approach for genome-wide association (WssGBLUP), which uses pedigree, genotypes, and phenotypes from genotyped and ungenotyped animals. The weighting strategy accounted for 1Mb moving SNP-windows along each of the 29 chromosomes in the reference genome. Genomic regions that explained more than 1% of the additive genetic variance were considered associated with CD resistance. A total of 13 windows located on six chromosomes were found to be associated with CD resistance in the NCCCWA population. Two windows, located at 59-60 Mb and 61-62 Mb on chromosome Omy17, explained 12% and 11.33% of the genetic variance for CD resistance, respectively. In the TLUM population, a total of 16 windows located on nine chromosomes were detected. Only three similar windows (located on two chromosomes) were detected in both populations. The results suggest that CD resistance has an oligogenic architecture, and the SNP windows found to be associated with CD are not informative enough for selection decisions across populations. In the next steps, we will assess strategies for genomic selection by predicting and comparing the accuracy of genomic evaluations generated using lower-density SNP panels and a panel composed solely from QTL-associated SNPs.

P0257: Aquaculture

Egg Transcripts Associated with Family Fertility in Rainbow Trout (*Oncorhynchus mykiss*)

Hao Ma[1], Kyle E. Martin[2], Doug Dixon II[2] and Gregory M. Weber[1], (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)Troutlodge, Inc., Sumner, WA

In order to identify molecular markers associated with rainbow trout egg quality, the RNAs of unfertilized eggs derived from 4 low and 10 high fertility families were sequenced using both poly(A) retention and rRNA removal kits. The results indicated that although the sequences derived from rRNA removal libraries identified about 20% more transcripts than that of poly(A) retention libraries, only one transcript was differentially expressed between the rRNA removal libraries, while 944 differentially expressed transcripts (DETs) were identified from poly(A) retention libraries under the same criteria. The 945 DETs were classified into 31 functional modules, of which 9 were related to ribosomes and 8 related to mitochondria. The other modules contained transcripts primarily involved in transcription, translation, cell division, apoptosis, and immunity. Small RNA sequencing of the low and high survival families identified 26 known and 38 novel differentially expressed miRNAs, which targeted 271 of the DETs. These results could be used to develop molecular markers to assess rainbow trout family fertility.

P0258: Aquaculture

Similar Effects of QTL Haplotypes for Bacterial Cold Water Disease Resistance across Two Generations in a Commercial Rainbow Trout Breeding Population

Yniv Palti[1], Sixin Liu[1], Roger L. Vallejo[1], Kyle E. Martin[2], Jason P. Evenhuis[1], Guangtu Gao[1], Gregory D. Wiens[1] and Timothy D. Leeds[1], (1)USDA-ARS-NCCCWA, Kearneysville, WV, (2)Troutlodge, Inc., Sumner, WA

Previously we have demonstrated that genomic selection (GS) for bacterial cold water disease (BCWD) resistance can double the accuracy of traditional pedigree-based selection in a commercial rainbow trout breeding population. The objective of this study was to evaluate the effectiveness of marker assisted selection (MAS). Application of MAS is more economical than GS as genotyping fewer SNPs has less cost. In addition, SNP haplotypes that are in tight linkage disequilibrium with the favorable QTL alleles can be used for direct selection of breeding animals eliminating the need for re-phenotyping and re-training of the GS or P-BLUP models in every generation. Here, we have confirmed that the favorable SNP haplotypes from three major QTL we have identified in the May 2013 nucleus breeding population of Troutlodge, Inc. were still strongly associated with improved BCWD resistance in the following May 2015 generation. Furthermore, using progeny testing data we have found that the combined predictive ability, or accuracy, of the favorable QTL haplotypes was consistent across two generations, and substantially better than the accuracy of the pedigree based BLUP predictions of genetic merit. Overall, our results indicate that MAS using a combination of favorable haplotypes from few large-effect QTL regions can rapidly and accurately identify elite candidate breeders for BCWD resistance in a commercial rainbow trout breeding population.

P0259: Aquaculture

Genome-Wide Identification of Antisense lncRNAs and their Expression and Genetic Polymorphism Associated with Susceptibility to *Flavobacterium psychrophilum* in Rainbow Trout

Ali R. Ali[1], Rafet Al-Tobasei[2] and Mohamed (Moh) Salem[1], (1)Middle Tennessee State University, Murfreesboro, TN, (2)University of Alabama at Birmingham, Birmingham, AL

Eukaryotic genomes encode several antisense long-noncoding RNAs (AS-lncRNAs) that regulate gene expression. Recently, we identified few AS-lncRNAs correlating with immune-related genes, however, a systematic genome-wide analysis of AS-lncRNAs in rainbow trout is lacking. To identify antisense transcripts, 134 RNA-Seq libraries from 5 different datasets were used. A total of 14,011 AS-lncRNA were identified genome-widely. 75.5% of AS-lncRNAs showed multiple exons compared to 36.5% of the intergenic lncRNAs. About 7% of AS-lncRNAs exhibited tissue-specific expression. RNA-Seq libraries from three genetic lines of rainbow trout with significant differences in survival rate following *F. psychrophilum* infection were analyzed to investigate role of the AS-lncRNAs during infection. 24 pairwise comparison between the different genetic lines, infectious status, and time points revealed 421 AS-lncRNA with differential expression (DE). In addition, 234 sense genes showed differential and strongly correlated expression with their AS-lncRNAs. The list of the DE sense/AS pairs includes immune-related genes. Of them, six inverse correlations were identified between DE AS-lncRNAs and coding counterparts such as genes encoding for CTP synthase and nuclear factor 1 C-type. Promoter analysis of sense/AS pairs revealed high representation of transcription factors binding sites that are critical to immune cell development/response including c-Jun, STATs, IRFs, PAX-5, and GATA3. Additionally, a 50K cSNP-chip identified 338 SNPs within sense/AS-lncRNAs overlapping regions. Those SNPs showed allelic imbalances between genetic lines and infectious status. Collectively, AS-lncRNAs represent an important layer of the molecular architecture of fish immunity, and could be used to develop biomarkers for genomic selection and genetic manipulation in aquaculture.

P0260: Aquaculture

Small RNAs Involvement in *Flavobacterium psychrophilum*-Rainbow Trout Host Pathogen Interactions

Pratima Chapagain[1], Ali R. Ali[1], Yniv Palti[2], Gregory D. Wiens[2] and Mohamed (Moh) Salem[1], (1)Middle Tennessee State University, Murfreesboro, TN, (2)USDA-ARS-NCCCWA, Kearneysville, WV

Gram-negative bacteria's outer membrane vesicles (OMV's) contain sRNAs and virulence factors that can be released from the bacterial surface during host-pathogen interaction. We hypothesized that sRNAs interact with host genes and impact their expression. To begin to test this hypothesis we first mapped all publicly available RNA-Seq reads from *F. psychrophilum* (Fp) onto the Fp genome (strain CSF-259). A total of 267 sRNAs were identified in the Fp genome. We then used computational analyses to identify 535 trout's immune-related genes possibly targeted by these sRNAs (free energy cut off value ≤-35.0kcal/mol).

Potential interaction of 10 sRNAs-trout immune genes were investigated in two rainbow trout genetic lines, ARS-Fp-R (resistant) and ARS-Fp-S (susceptible), created by selective breeding and showing significant variation in the survival rate after 21 days of exposure to the *Fp.* qPCR was used to determine the reciprocal expression of the sRNA and their targets in whole-body lysates after 5 days of infection. B-actin and Prolyl-tRNA synthetase were used as reference genes for trout mRNA and Fp-sRNA measurements, respectively.

Interestingly, three immune-related genes; macrophage migration inhibitory factor (MIF), autophagy related protein-9A and mitogen activated protein kinase-7 (MAPK) were downregulated in the susceptible compared to resistant fish, whereas, three sRNAs targeting those genes were reciprocally upregulated (p-value<0.05). Additionally, two other sRNAs showed significant differences in resistant versus susceptible fish without significant differences in trout mRNA abundance. Further research is required to confirm the presence of these sRNA within vesicles and directly measure the effect of OMV's on the trout immune response.

P0261: Aquaculture

The Genome of the Coho Salmon

Eric B. Rondeau[1], Jong S. Leong[1], David R. Minkley[1], Kris A. Christensen[2], Cody A. Despins[1], Anita Mueller[1], Robert H. Devlin[2], Ruth E. Withler[3], Terry Beacham[4], Kerry A. Naish[5], Jose Manuel Yanez[6], Roberto Neira[7], Louis Bernatchez[8], William S. Davidson[9] and Ben F. Koop[1], (1)University of Victoria, Victoria, BC, Canada, (2)Fisheries and Oceans Canada, Centre for Aquaculture and Environmental Research, West Vancouver, BC, Canada, (3)Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada, (4)Molecular Genetics Lab, Pacific Biological Station, Nanaimo, BC, Canada, (5)University of Washington, Seattle, WA, (6)Aquainnovo, Puerto Montt, Chile, (7)University of Chile, Santiago, Chile, (8)Université Laval, Québec, QC, Canada, (9)Simon Fraser University, Molecular Biology and Biochemistry Department, Burnaby, BC, Canada

Coho (*Oncorhynchus kisutch*) is a culturally and economically important salmon species that spawn in rivers that flow to the Northern Pacific Ocean. Wild coho salmon stocks have seen significant declines over the past quarter century, with few signs of recovery. An emerging species in aquaculture in its native range, the species is an established component of the aquaculture industry in Chile. To assist in wild stock management and to further develop hatchery and aquaculture stocks, genomic tools for coho salmon are being developed. Such efforts are complicated by the salmonid whole-genome duplication event approximately 80-100 MYA which resulted in a large, highly similar, pseudo-tetraploid genome. In this work, we present a chromosome-level assembly of the Coho salmon genome. This first release of the assembly has a contig N50 of 58 kbp, a scaffold N50 of 1,266 kbp and a longest scaffold of 15,030,138 bp. The assembly was anchored to a previously published linkage map, with 71.1% of assembled bases assigned to a chromosome. A transcriptome was generated from fifteen tissue-specific libraries for genome annotation and to generate a gene expression atlas. Sixty coho sourced from twelve locations throughout the North American range have been re-sequenced to evaluate variation and to develop a high-density SNP array for consistent genotyping in further populations. The Coho salmon genome and the development of genomic resources will aid in management and conservation of wild populations, the analysis of hatchery efficacy and improvements in aquaculture production.

P0262: Aquaculture

Sequencing and Assembling the Genomes of the Chinook, Sockeye, Pink, and Chum Salmon

Kris A. Christensen[1], Jong S. Leong[2], Dionne Sakhrani[1], David R. Minkley[2], Ruth E. Withler[3], Eric B. Rondeau[2], Ben F. Koop[2] and Robert H. Devlin[1], (1)Fisheries and Oceans Canada, Centre for Aquaculture and Environmental Research, West Vancouver, BC, Canada, (2)University of Victoria, Victoria, BC, Canada, (3)Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, Canada

To unify genomic resources between studies and effectively compare data between species, there is no better resource than a genome sequence. In order to capitalize on the already available genomic resources, including genome sequences from other salmonids, we are sequencing and assembling the genomes from Chinook, sockeye, pink, and chum salmon. The poster/presentation will outline the progress and future directions of this initiative, and includes information on how others may get involved.

P0263: Aquaculture

Using the Medium Density Panels to Increase the Feasibility of Imputation-Based Genomic Selection in Atlantic Salmon

Tiago Hori[1], America Fujimoto[2], John Buchanan[3] and Jason Stannard[3], (1)The Center for Aquaculture Technologies, Souris, PE, Canada, (2)The Center for Aquaculture Technologies, San Diego, CA, (3)Center for Aquaculture Technologies, San Diego, CA

Genomic selection has the potential to revolutionize breeding in aquaculture by increasing accuracy of selection and per generation genetic again, which in turn can lead to significant increases in profitability. The wide-spread application of genomic selection in Aquaculture has been hindered by the elevated cost of genotyping large training populations and sets of candidate broodstock in every production cycle. Even if training populations are not genotyped every cycle, the costs are still prohibitive and there is ample evidence that there is significant loss in accuracy if the prediction model is not updated (re-trained) every generation. Genotype imputation represents a powerful tool to reduce genotyping costs in genomic selection applications. However, given the size of the salmon genome, the number of SNPs required for genoytping and accurate imputation has historically been cost-prohibitive or beyond the capabilities of many genotyping platforms. In this presentation, we highlight present the use of medium-density panel panel that can be used to genotype large numbers of individuals at an effective cost. Simulation data will also be presented to demonstrate the imputation power of a 3,000 SNP panel and its application on various genomic selection schemes.

P0264: Aquaculture

Identification of Differentially Expressed Mirnas between Resistant and Susceptible Atlantic Salmon (*Salmo salar*) Fry Challenged with Infectious Pancreatic Necrosis Virus

Nardos Tesfaye Woldemariam, Oslo and Akershus University college of applied sciences, Oslo, Norway, Bjørn Høyheim, Department of Basic Sciences and Aquatic Medicine, Faculty of Veterinary Medicine and Biosciences, Norwegian University of Life Sciences, Oslo, Norway, Oleg Agafonov, Bioinformatics Core Facility, IFI, University of Oslo, Oslo, Norway, Ross D. Houston, The Roslin Institute and R(D)SVS, Edinburgh, United Kingdom, John B. Taggart, Institute of Aquaculture, University of Stirling, Stirling, United Kingdom and Rune Andreassen, Oslo and Akershus University College of Applied Sciences, Oslo, Norway

MicroRNAs (miRNAs) control multiple biological processes including the innate immune responses by negative post-transcriptional regulation of gene expression. In this study, we aimed to identify miRNAs responding to *infectious pancreatic necrosis virus* (IPNV) infection. Experimental samples comprised Atlantic salmon fed fry that were classed as either 'resistant' or 'susceptible' according to genotype at the major quantitative trait locus controlling resistance to infectious pancreatic necrosis. The fry samples were collected at three different time points (1, 7 & 20 days post challenge (poc)) to identify early to late responding miRNAs. *In silico* predictions of target genes were carried out to determine the putative function of responding miRNAs in immune gene networks.

RNA extracted from whole fry (n=96) were sequenced using the Illumina small RNA sequencing protocol, providing results from 96 independently sequenced libraries. DESeq2 was used to identify differentially expressed miRNAs (DE miRNAs) while RNAhybrid was applied for target gene predictions.

Thirty-six different mature miRNAs belonging to 27 families were identified as responding to IPNV challenge. There were twelve miRNAs with differential expression at 1 day poc, while there were 13 and 23 at 7 and 20 days poc, respectively. A significant difference between resistant and susceptible individuals was observed for 21 miRNAs (from 13 families) at the final time point only. Based on the direction of the expression changes and the target genes predicted by *in silico* analysis we discuss the putative function of the miRNAs in the immune response to this viral infection in Atlantic salmon. Their predicted functions need to be validated and further studied in functional assays to fully understand their roles in immune homeostasis.

P0265: Aquaculture

Genome Wide Association Analysis for Resistance to the Causal Agent of Bacterial Kidney Disease in a North American Commercial Atlantic Salmon

Melissa K. Holborn[1], Keng P. Ang[2], J.a.K. Elliott[3], Frank Powell[3] and Elizabeth G. Boulding[1], (1)University of Guelph, Guelph, ON, Canada, (2)Cooke Aquaculture, Saint John, NB, Canada, (3)Cooke Aquaculture, Oak Haven, NB, Canada

Bacterial kidney disease (BKD), caused by the pathogen Renibacterium salmoninarum, is economically costly to the Atlantic salmon farming industry. BKD mortality of infected fish may not occur for one to two years after infection, after a substantial monetary investment. Our goal was to identify associations between single nucleotide polymorphisms (SNPs) and resistance to BKD in the Saint John River strain of Atlantic salmon that could be used to select for resistant fish. 652 fish from 63 families were experimentally infected with R. salmoninarum by intraperitoneal injection. Fish were held in controlled tanks and daily mortalities were recorded. After 102 days, all surviving fish were euthanized and sampled. 576 fish were selectively genotyped on a custom 50K SNP chip designed for North American Atlantic salmon. 508 fish and 44,346 SNPs passed quality control. Genome wide association analysis was performed using the GenABEL package. A mixed model approach using a polygenic model was used to account for familial relatedness and population structure. A single SNP was found to have chromosome wide significance after Bonferroni correction when analyzing survival as the BKD resistance trait. Two SNPs were found to have chromosome wide significance after correction when analyzing time to death and the BKD resistance trait. These 3 SNPs explained a small percentage of the phenotypic variance and were located on different chromosomes, 4.0%, 3.8%, and 3.6% respectively. The results of these two association analyses indicating that BKD resistance, both as overall survival and time to death, has a polygenic trait architecture.

P0266: Aquaculture

Functional Genomic Basis of Host Resistance to Sea Lice in Atlantic Salmon

Ross D. Houston[1], Diego Robledo[1], Alejandro P. Gutierrez[1], Agustin Barria[2] and José M Yáñez[2], (1)The Roslin Institute and R(D)SVS, Edinburgh, United Kingdom, (2)University of Chile, Santiago, Chile

Sea lice are parasitic copepods that cause large economic losses to salmon aquaculture worldwide. Alternative methods to control this parasite, such as selection for host resistance, are increasingly important. Insight into the host-parasite interaction and mechanisms of host resistance can lead to improvements in selective breeding for resistance, and potentially novel treatment targets. The aims of this study were to characterise the functional genomic basis of host resistance to lice, and to identify potential functional polymorphisms underlying resistance. To achieve this, challenge experiments were performed on a population of salmon from a Chilean breeding program, from which a large genome-wide association study was performed using a SNP array. In addition, salmon from resistant and susceptible families were compared using RNASeq of attachment sites and healthy skin, and using whole genome sequencing. Analyses of the gene expression signature of host resistance revealed genes and pathways associated with resistance, and these results were cross referenced with the GWAS and WGS data to identify candidate functional resistance genes and polymorphisms. These results improve our understanding of host response to lice in salmon, and highlight potential functional genomic variants that could be used to enhance genomic selection for host resistance to lice in salmon breeding programs to help tackle this major disease problem.

P0267: Aquaculture

Sequencing Albacore (Thunnus alalunga) Genome to Identify Sex Specific Markers

Arun S. Seetharam[1], **Catherine Purcell**[2], Rick Masonbrink[1], Owyn Snodgrass[2], Steve Teo[2], John Hyde[2] and Andrew J. Severin[1], (1)Iowa State University, Ames, IA, (2)Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, CA

While sex is one of the most basic biological characteristics of an organism, it can also be one of the most challenging to identify, particularly for marine fish. Many marine fish lack morphological characteristics that distinguish sex, and even with dissection, sex may be difficult to determine in immature individuals. Sex identification is a high priority in the North Pacific Albacore Tuna (*T. alalunga*), as early identification can facilitate aquaculture stock assessment. The development of genetic markers for sex would increase identification speed and provide a non-destructive alternative to dissection (e.g., fin clip samples from tagged fish). To improve the odds of identifying a sex- specific region in Albacore, this study used a next generation sequencing approach using paired-end and mate-paired data to assemble a draft genome for *T. alalunga*. To identify genomic regions associated with sex, lower coverage sequencing was conducted on additional male and female specimens (ten of each sex). A preliminary assembly was generated from the pooled data of several individuals using Platanus (v1.2.4). This assembly was further processed with Redundans (v0.11) to reduce heterozygosity and collapse haplotypes. The total size of assembled scaffolds is approximately 80 % (724.4Mb) of the expected genome size, with an N50 of 167 Kb and L50 of 1,208bp. A total of 3,130 (68.3% %) BUSCO genes (actinopterygii_odb9 profile) were found in the assembly. BRAKER was then used to predict 39,590 genes using publicly available RNAseq data. The next steps in this project are to use a genome-wide association study to identify the sex determining region.

P0268: Aquaculture

Chinese Tongue Sole: Genome to Breeding

Songlin Chen[1,2], Zhongkai Cui[1], Changwei Shao[1], Yang Liu[1], Yangzhen Li[1], Na Wang[1], Qian Zhou[1], Feng Liu[1], Sheng Lu[1], Wenteng Xu[1] and Manfred Schartl[3], (1)Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, (2)Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, (3)University of Wuerzburg, Wuerzburg, Germany

Chinese tongue sole (*Cynoglossus semilaevis*) is an economically important marine flatfish with a ZW sex chromosome system. Females are 2-4 times bigger than males at same age. We assembled whole genomes of tongue sole. *dmrt1* was determined to be candidate for sex-determining gene. It is Z chromosome-linked and male specifically expressed. Genome editing by TALEN reveals *dmrt1* as an essential male sex-determining gene linked to sexual dimorphism in Chinese tongue sole. The *dmrt1* mutated males grow faster as females.

By comparing female and male genome, some sex-linked SSR markers were found and used to develop molecular technique for genetic sex identification of ZZ, ZW and WW fish. We found that phenotypic female accounts for only 10-30% in cultured populations, and more than 90% of pseudo-male offspring sex-reverted into a pseudo-male. Whole-genome methylation sequencing revealed that all second-generation pseudomales had inherited the Z chromosome from their sex-reversed fathers and retained the paternal methylation pattern. Finally a molecular marker-assisted sex control method was developed to increase phenotypic female ratio.

In addition, diseases often resulted in huge economic loss in tongue sole aquaculture, genomic selection is a promising technique for breeding disease-resistant fish. Based on many families developed during last 5 years, we constructed reference population and carried out genome resequencing, predicted GEBV for disease-resistant trait for the reference population in the tongue sole. Finally we have established genomic selection method and applied it to breeding of tongue sole with enhanced disease resistance to pathogenic bacteria, *V. harveyi.*

P0269: Aquaculture

The Role of Z-Chromosome-Specific Ubiquitin Ligase Genes in Spermatogenesis of Chinese Tongue Sole

Wenteng Xu[1], Hailong Li[1], Liang Meng[1] and Songlin Chen[1,2], (1)Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, (2)Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

E3 ubiquitin ligases are a large gene family that plays a diversity of roles in spermatogenesis. Based on the genomic sequencing data, over 200 E3s have been identified across the whole genome of Chinese tongue sole (*Cynoglossus semilaevis*) and 6 Z-chromosome-specific candidates (*neurl3*, *trim25-like*, *rnf34-like*, *dtx3l-like*, *dtx1*, *rchy1*) were particularly focused as their sexually differential transcription patterns in male, female, pseudomale. Among them, *neurl3* was further selected for functional characterization. Firstly, we found that *neurl3* transcription was male-biased and coincidentally overlapped with threshold of spermatogenesis. *neurl3* translation was predominant in male testis and immunohistochemistry revealed its localization in germ cells. Then the *in vivo* knockdown of *neurl3* was conducted by siRNA-mediated interference, which resulted in increased transcription of spermatogenesis-related genes. Moreover, the levels of *neurl3* transcription and testis

protein ubiquitination were found to be closely correlated. Based on these findings, we speculate that *neurl3* modulates testis protein ubiquitination in a dosage-dependent manner and that this influences spermatogenesis.

P0270: Aquaculture

Physical Mapping of Genome in Yellowtail (*Seriola quinqueradiata*) and Population Structure in the Pacific Ocean

Akiyuki Ozaki[1], Takashi Isikawa[1], Jun-ya Aoki[1], Junya Kawase[2], Takurou Shinkai[2], Junko Tabata[1], Kazunori Yoshida[1], Tsutomu Noda[3] and Kazuo Araki[1], (1)National Research Institute of Aquaculture - National Research Development Agency, Mie, Japan, (2)Marine Biological Science, Graduate School of Bioresources, Mie University, Mie, Japan, (3)Seikai National Fisheries Research Institute, National Research Development Agency, Nagasaki, Japan

Expectation of aquaculture research is increasing in response to predictions of depletion of natural aquatic resources. Yellowtail is distributed in Pacific coastal areas of Japan, and is an important aquaculture fish “Hamachi” in Japan. For genetics breeding by marker-assisted selection or genomic selection, it is important to construct a reference genome sequence by identification of single nucleotide polymorphisms (SNPs) as well as to identify functional genes for useful biological traits. And high-density linkage maps generated by SNPs data have proven to be crucial for the accurate assembly of scaffolds and contigs for whole-genome sequencing.

We performed *de novo* sequencing of one yellowtail genome. The final de novo genome assembly was 639.3 M bp, with 384 scaffolds, and an N50 scaffold size of 5.6 Mbp. It lined up genome sequence on the radiation hybrid physical map and was covered by 188 long scaffolds. We are developing re-sequence of each of 16 individuals to detect SNPs, and to reorder the 24 linkage groups of the physical sequence and radiation hybrid maps.

The SNPs were used to check population fragmentation in Pacific coastal areas of Japan. To confirm the effectiveness of the SNPs, 96 neutral autosomal SNPs in noncoding regions were selected from each of the linkage groups. And we carried out principal components analysis of five populations of yellowtail using E-gene strata. As such it was difficult to distinguish genetically separate subdivisions. This means that a high genetic variance occurs in the wild yellowtail population, implying potential for use for genetic improvement.

P0271: Aquaculture

Genome Assembly and Annotation for Almaco Jack (Seriola rivoliana)

Arun S. Seetharam[1], Catherine Purcell[2], Rick Masonbrink[1], Sivanandan Chudalayandi[1], Mark Drawbridge[3], John Hyde[2] and Andrew J. Severin[1], (1)Iowa State University, Ames, IA, (2)Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, CA, (3)Hubbs-SeaWorld Research Institute, San Diego, CA

Amberjacks are a fish of interest to the aquaculture industry, as they comprise more than a billion dollars of the sashimi industry. Amberjacks are a colloquial term for a group of *Seriola* species (*S. dumerili, S. dorsalis, S. rivoliana, S. quinqueradiata*). *S. rivoliana* is a common target for aquaculture in the United States, as wild distributions are found only in Hawaii and South America and are not commercially harvested from the wild due to ciguatera poisoning and parasitic worms. Successful breeding in aquaculture can prevent these conditions. Thus, to facilitate aquaculture improvements genomic resources are needed to identify pertinent *Seriola* genes. Here we report a genome assembly for *S. rivoliana* using tissue from four *S. rivoliana* individuals (two males and two females). Paired-end and mate pair libraries were prepared and sequenced using Illumina HiSeq 3000 sequencer (50-60X per sample), while long reads were generated using a PacBio RS-II instrument (20 SMRTcells, ~20X). The initial assembly was generated from a combination of short and long reads with MaSuRCA (v 3.2.2) on XSEDE computing resources. Additional scaffolding was performed using Dovetail Chicago libraries and HiRise program. The total size of assembled scaffolds is approximately 98 % (667.1Mb) of the expected genome size, with an N50 of 9.5 Mb and an L50 of 23. The genome is highly complete, demonstrating a 94.6 % (4,338 genes, actinopterygii_odb9 profile) BUSCO complete score. The next steps are to predict genes, which will be performed using BRAKER using the RNAseq data from *S. dorsalis*.

P0272: Aquaculture

A Diagnostic SNP Panel for Assessing Purity and Hybridization of the Black Basses (*Micropterus spp.*)

Eric Peatman[1], Wilawan Thongda[2], Honggang Zhao[2] and Lauren Davis[2], (1)AUBURN UNIVERSITY, Auburn, AL, (2)AUBURN UNIVERSITY, AUBURN, AL

Black basses (*Micropterus* spp.) are among the most important recreational sportfish in the United States. Depending on the species, they are of additional interest for aquaculture, conservation, and/or ecological monitoring. *Micropterus* species can readily hybridize, particularly as stocking events bring species lacking reproductive isolating mechanisms into contact with one another. Several black bass species across the Southeastern United States are highly vulnerable to swamping events via hybridization with introduced sister species. Rapid identification of pure and hybridized individuals is critical in state agency hatcheries tasked with restorative stocking and in monitoring the health of riverine populations. To facilitate these assessments, we developed panels of fixed/diagnostic SNP markers for black bass. Following genotyping-by-sequencing, we validated and extended a panel of 64 markers capable of differentiating 13 species and 2 undescribed types of black bass. We have tested these markers on 1304 samples to-date. Case studies illustrating the informativeness of the markers, taken from analyses of natural populations of bass in Alabama, Georgia, and Texas will be presented.

P0273: Aquaculture

Construction of High-Density Genetic Linkage Maps and QTL Mapping in the Golden Pompano

Jian Luo, State Key Lab of Marine Resource Utilization in South China Sea, Hainan University, Haikou, China and Guoqing Zhang, Xinhui Zhang, Hengzhen Ye, Shoujia Jiang, Hui Yu, Jia Li, Qiong Shi, Guohua Chen, Zhi Zhou, Jian Luo, Xinxin You

The golden pompano (*Trachinotus blochii*) is an essential marine fish species of South China. In this study, a genetic map containing 24 linkage groups was compared with 102 F_1 generation individuals (full-sib). A total of 23,558 single-nucleotide polymorphisms (SNPs) have been identified via restriction-site associated DNA sequencing (RAD-seq). Among these, 12,358 high-quality SNPs were utilized to construct a linkage map and consequently, we present the first reported genetic map of *T. blochii*. The total lengths of female, male and sex average linkage maps were 2,559.12 cM, 2,827.48 cM and 3,810.03 cM, respectively. These linkage maps were built via the application of quantitative

trait loci (QTL) tags, which are related to the size and growth of *T. blochii*. We successfully detected 23 QTL of growth-related traits, including 4 QTL for body length, 5 QTL for body weight, and 14 QTL for body height. This newly developed, QTL facilitating, and high-resolution genetic map represents preliminary progress toward linking the traits of key conservation interest to DNA regions in *T. blochii.* Our research will provide guidance for the molecular breeding of the golden pompano.

P0274: Aquaculture

Genome-Wide Identification and Characterization of Long Noncoding RNAs in Tilapia *Oreochromis niloticus*

Walaa Shaalan[1], **Ali R. Ali**[2], Nassr Allah Abd El-Hameid[1], Sabry El-Serafy[1] and Mohamed (Moh) Salem[2], (1)Department of Zoology, Faculty of Science, Benha University, Benha, Egypt, (2)Middle Tennessee State University, Murfreesboro, TN

Long noncoding RNAs (lncRNAs) regulates various biological processes including fish growth and disease resistance. In this study, RNA sequencing reads from 12 tissues of tilapia were mapped to a genome reference and assembled into 42,200 transcripts using HISAT2/StringTie suite. In total, 16,026 lncRNAs were identified after the filtration of transcripts with protein-coding potential. Of them, 6,118 transcripts were categorized as antisense lncRNAs (AS-lncRNAs), transcribed from the opposite strand of protein-coding loci. 69.5% and 16.6% of lincRNAs and AS-lncRNAs were multi exonic, respectively. Investigation of lncRNA tissue specificity revealed 2,066 and 1,010 tissue-specific lincRNA and AS-lncRNA transcripts, respectively. Remarkably, ovary and testis showed the largest number of tissue specific lncRNA transcripts; 678 and 837, respectively. This study represents a comprehensive genome-wide analysis of tilapia lincRNAs/AS-lncRNAs and provides invaluable genomic resources for future functional genomics studies in tilapia.

P0275: Aquaculture

Adapting Stickleback As a Model for Host-Microbe Interaction Studies

Kathryn CA Milligan-Myhre, Emily Lescak, Kelly Ireland, Ryan Lucas, Kat O'Brien and Kenneth Sparks, University of Alaska Anchorage, Anchorage, AK

Threespine stickleback (*Gasterosteus aculeatus)* fish have recently been adapted to study the evolution and relationship between hosts and their microbiota. Microbiota are communities of microbes associated with the host. These communities often have thousands of different species, each with different roles, most of which are acquired from the environment and food. We use wild-caught and lab raised stickleback to determine: 1. Do hosts who have evolved in different microbial environments select for different microbes? 2. Does the immune response to microbes affect the selection for microbes that will become part of the microbiota? 3. Does the host genetic background influence the ability of the host to maintain a stable microbiota when challenged with environmental or antibiotic treatments? We have found: 1. Two stickleback populations have microbiota that are different from each other. We are broadening this to six populations, and preparing a common garden experiment to determine the role the host genetic background plays in selection of the microbiota. 2. One anadromous and one freshwater population have different inflammatory responses to microbes. We are developing new immune assays and broadening this to three populations each of anadromous and freshwater fish. 3. Fish from different genetic backgrounds have different developmental trajectories when challenged with antibiotics. We are designing a common garden experiment with long-term antibiotic and environmental contaminant exposure to determine long-term effects of microbiota challenge on host development. Most of these experiments require new aquaculture facilities, so we are developing those as well.

P0276: Aquaculture

Shrimp Diseases and the Virulence Factor Content of Aquaculture *V. parahaemolyticus* Isolates

Sarah Mae Penir[1,2] and Cynthia P. Saloma[1,3], (1)Philippine Genome Center, University of the Philippines, Quezon City, Philippines, (2)PCARI Shared Genomics Core Laboratory, University of the Philippines, Quezon City, Philippines, (3)National Institute of Molecular Biology and Biotechnology, University of the Philippines, Quezon City, Philippines

Shrimp/prawn is among the major export commodities of the Philippines. One well-documented pathogen of penaeid shrimp is *Vibrio parahaemolyticus*, a Gram-negative, halophilic bacteria with a single polar flagellum. Most environmental or seafood-isolated strains of *V. parahaemolyticus* are not pathogenic, but some strains become capable of causing diseases due to the presence of certain virulence factors. In order to gain insight on the pathogenicity of the bacterium, detection of the virulence factor sequences obtained from the Virulence Factor Database (VFDB) was performed on the raw reads of 137 Philippine shrimp *V. parahaemolyticus* isolates through a mapping approach. The virulence factor content of the Philippine isolates was compared with the whole-genome sequences of 10 publicly available shrimp-associated isolates. Principal component analysis (PCA) on the virulence factor content clearly distinguished the Philippine isolates from those originating in other countries. A thorough understanding of the pathogenicity of *V. parahaemolyticus* through its virulence factors can pave the way for the establishment of effective shrimp disease management interventions, such as the development of fast and affordable detection tools, which can preclude further severe losses in shrimp production.

P0277: Aquaculture

Advanced Black Tiger Prawn Breeding using DNA Markers and High-Throughput Phenomics from Commercial Ponds

Mehar S Khatkar[1], Greg J Coman[2], Nick M Wade[2], Kyall R Zenger[3], Casey Bameja[3], Jarrod Guppy[3], Andrew Foote[2], Melony J Sellars[2], Peter C Thomson[4], Herman W Raadsma[1] and Dean R Jerry[3], (1)Sydney School of Veterinary Science, The University of Sydney, and ARC Research Hub for Advanced Prawn Breeding, Camden, Australia, (2)Aquaculture Program, CSIRO Agriculture and Food, and ARC Research Hub for Advanced Prawn Breeding, St. Lucia, Australia, (3)Centre for Sustainable Tropical Fisheries and Aquaculture, James Cook University and ARC Research Hub for Advanced Prawn Breeding, Townsville, Australia, (4)School of Life and Environmental Sciences, The University of Sydney, and ARC Research Hub for Advanced Prawn Breeding, Camden, Australia

Limited information is available on the results of practical implementation of genomic selection (GS) in aquaculture species, which have mainly been restricted to the use of simulated data. The feasibility of GS schemes in aquaculture depends on the availability of cost-effective genomic resources and extensive production trait records collected on the reference populations. Here we describe the generation of a resource

or training population of black tiger prawn (*Penaeus monodon*) for genetic analyses and gene discovery, and its utility in the development of an advanced breeding program for long-term genetic improvement i.e. implementing genomic selection, as part of the ARC Industrial Transformation Research Hub for Advanced Prawn Breeding, an Australian organisation consisting of a partnership among universities, CSIRO and the private sector. We have applied low-cost genotyping, DNA pooling and phenomics for recording shrimp performance under commercial pond environments. Our results indicate that accurate estimation of genomic relationships can be achieved with relatively low-density SNP panels (1000-3000 SNPs) and are comparable to those estimated from medium-to high density SNP panels (50,000+ SNPs). We have used high throughput phenomics using digital imaging and near infra red (NIR) spectroscopy, and machine learning algorithms for recording of quantitative phenotypic information on a large number of animals (>25,000) on the farm. The collection of disease resistance data is critical for a successful breeding program in black tiger prawn. The practical utility of pooled genotype data on pre-and post-challenge of large mixed-family progeny cohorts exposed to standardised disease challenge and ascertainment of survival statistics will be discussed.

P0278: Aquaculture

High-Density ddRAD Linkage Map and Temperature, Salinity and Alkalinity Resistances-Related QTL Mapping of Pacific White Shrimp (*Litopenaeus vannamei*)

Ting Chen, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

Pacific white shrimp (*Litopenaeus vannamei*) is the dominant shrimp species produced in global seafood mariculture. However, the farming of *L. vannamei* seriously suffers from some environmental factors such as temperature, salinity and alkalinity. The whole genome sequencing of *L. vannamei* has not been completed, thus the molecular marker-related breeding needs to be carried out by other ways. Double-digest restriction-site associated DNA sequencing (ddRAD-seq) is a powerful and inexpensive approach to developing numerous single nucleotide polymorphism (SNP) markers and constructing a high-density genetic map. In this study, we constructed a ddRAD-based genetic map of *L. vannamei*. Based on the high-density linkage map, several QTLs for low-temperature, low salinity and high alkalinity resistances were detected. Some resistance-related SNP were located to functional genes by the comparison of the 2 generation and 3 generation transcriptomes. The genome survey and high-density genetic map construction may provide genomic and genetic resources for the breeding of *L. vannamei*.

P0279: Aquaculture

Assessing the Potential for Inhibiting Viral Infection in Whiteleg Shrimp *Litopenaeus vannamei* using FANA Antisense Oligonucleotide Technology

E HU, Center for Aquaculture Technologies, Inc, san diego, CA, Veenu Aishwarya, AUM LifeTech, Inc., Philadelphia, PA, **Tiago Hori**, The Center for Aquaculture Technologies, Souris, PE, Canada and John Buchanan, Center for Aquaculture Technologies, San Diego, CA

RNAi has conventionally been used as an approach for gene knockdown in shrimp antiviral research, including protection of shrimp from viral infection. That said, RNAi has some caveats including delivery issues, stability and off target effects. Here we report a novel approach using next generation FANA antisense oligonucleotides (FANA ASOs) to reduce protein expression in *Litopenaeus vannamei* with a potential to inhibit virus replication in shrimp. Single stranded FANAs ASOs have attractive characteristics compared to conventional dsRNAs (using RNAi approaches) as they can be self-delivered, are highly stable, nontoxic and have high affinity and specificity to the target RNA. In this study we investigated the ability of FANA ASOs to bypass the RNAi pathway and still have effective and efficient protein reduction.
The objectives of this study were to: 1) establish an assay for measuring gene silencing using clotting protein (CP); 2) evaluate FANA ASOs targeting CP mRNA via steric blocking, and 3) compare the efficacy of FANAs and dsRNAs for inducing the desired hemolymph clotting phenotype. With conventional RNAi approach, we observed a reduction in CP mRNA and hemolymph clotting. However, with steric blocking FANA ASOs we observed reduced hemolymph clotting but no reduction on mRNA level. This demonstrated that FANA oligos did not induce mRNA cleavage of CP, instead, they blocked translation of CP which led to lack of CP during wound healing. Considering the advantages of FANAs, this unique approach makes FANA ASOs potentially an effective alternative for targeting viral protein synthesis and inhibiting pathogenic virus replication.

P0280: Aquaculture

The Eastern Oyster Genome: A Resource for Comparative Genomics in Shellfish Aquaculture Species

Wesley Warren[1], Ximing Guo[2], Dina A. Proestou[3], Patrick Minx[1], Chad Tomlinson[1] and **Marta Gomez-Chiarri**[4], (1)McDonnell Genome Institute at Washington University, St. Louis, MO, (2)Rutgers University, Port Norris, NJ, (3)USDA ARS National Cold Water Marine Aquaculture, Kingston, RI, (4)University of Rhode Island, Kingston, RI

Oyster aquaculture is an important sector of world food production. As such, it is imperative to develop a high quality reference genome for the eastern oyster, *Crassostrea virginica,* to assist in the elucidation of the genomic basis of commercially and ecologically important traits. We predict that the use of all types of sequence variants (e.g. genetic, gene expression, methylation) will expand the association of allelic states with trait variance. Genetic, gene expression, and methylation studies depend on the use of accurate assembled consensus base sequences. Major challenges to the assembly of molluscan genomes are the high rate of polymorphism and the unique repeat architecture present in these species. To partially address these challenges, DNA from an eastern oyster produced by gynogenesis within an inbred line was used for single molecule real time sequencing of the genome (~87x coverage). All reads were assembled using FALCON to a total size of 684Mb with an N50 contig length of 1.9Mb. The predicted genome size is reported to be 578-675Mb. Total repetitive elements were estimated to be ~36% using WindowMasker. A majority of sequences (>99%) were scaffolded into the known number of 10 chromosomes using a HiC proximity map then aligned to an eastern oyster genetic linkage map to confirm correct association. Gene annotation using the automated NCBI pipeline predicts the presence of 34,596 protein coding genes and 4,230 non-coding. This reference resource should facilitate discovery of a wealth of information about traits important to oyster health and development.

P0281: Aquaculture

Differential Expression of Apoptosis Pathway Gene Families in Response to Immune Challenge in *Crassostrea gigas* and *Crassostrea virginica*

Erin M. Roberts and Marta Gomez-Chiarri, University of Rhode Island, Kingston, RI

The eastern oyster, *Crassostrea virginica,* and Pacific oyster, *C. gigas*, are affected by disease outbreaks which threaten industry sustainability and ecosystem function. Oysters rely on a complex innate immune system characterized by several significantly expanded innate immune gene families. The genetically controlled pathway of programmed cell death, apoptosis, plays significant roles in immunity. Expansion of gene families involved in apoptosis is confirmed in *C. gigas* and supported by *C. virginica de novo* transcriptomic studies. The role of apoptosis in disease resistance in these species is unknown. This study utilized the recently available *C. virginica* genome to perform differential gene expression analysis on publically available transcriptomes of oysters challenged with a variety of stimuli, including Oyster Herpesvirus OsHV-1, several *Vibrio* spp., *Micrococcus luteus*, *Alliroseovarius crassostreae* CV919-312 (cause of Roseovarius Oyster Disease), and the probiotic bacterium *Bacillus pumilus* RI-695. Intra-species analysis reveals unique apoptosis pathway responses between challenges, and changes in functionally enriched pathways identified by Gene Set Enrichment Analysis. Bacterial challenge significantly enriched metabolic processes and chromosomal maintenance genes. Viral challenge significantly enriched signal transduction and apoptosis inhibitors, with Inhibitor of Apoptosis 2 (IAP2) and suppressor of cytokine signaling 2 most significantly differentially expressed. Transcripts from the expanded families IAP and GTPase of the Immune Associated Proteins (GIMAP) show distinct patterns of expression depending on the nature of the immune challenge. Cross-species comparison of apoptosis pathway expression indicates promising genetic targets for pathway manipulation under disease challenge and potential candidates for disease resistance markers.

P0282: Aquaculture

Expansion of Stress-Associated Genes in the Genome of the Sydney Rock Oyster, *Saccostrea glomerata*

Daniel Powell[1], Wayne O'Connor[2], David Raftos[3] and Abigail Elizur[1], (1)University of the Sunshine Coast, Sippy Downs, QLD, Australia, (2)NSW Department of Industry, Department of Primary Industries, Taylors Beach, Australia, (3)Macquarie University, North Ryde, Australia

Oysters perform important roles in estuarine ecosystems and are of substantial economic value to fisheries and aquaculture worldwide. Contending with disease and environmental stress are considerable challenges to oyster culture. The Sydney rock oyster, *Saccostrea glomerata*, is an iconic and commercially important species of edible oyster in Australia contributing a production value of over AUD$36 million for the state of New South Wales in 2015/2016. In Australia, wild populations of *S. glomerata* have been affected from the introduction of the faster growing *Crassostrea gigas*, which can rapidly overgrow and displace the native oyster at the low to mid-intertidal zones. At higher intertidal areas, however, *S. glomerata* can persist due to its higher tolerance to the tougher abiotic conditions. A draft genome for this species has been recently developed enabling detailed comparative studies. To better understand the genome-wide similarities among molluscs and other animals, an examination of the distribution of 8,629 Pfam domains across a diverse set of 26 metazoan genomes was performed. This analysis identified several significantly enriched Pfam domains in Saccostrea that were considered to represent gene family expansions. Among the 29,738 predicted protein-coding genes in the oyster genome, notable expansions of gene families associated with stress response such as heat shock protein HSP70, universal stress proteins (USPs) and inhibitor of apoptosis proteins (IAPs) were observed. This analysis offers insight into the genomic background of this oyster's resilience to abiotic stress.

P0283: Aquaculture

Draft Genome of the Hong Kong Oyster C*rassostrea hongkongensis*

Yang Zhang, south china sea institute of oceanology, Guangzhou, China

Hong Kong oyster *Crassostrea hongkongensis,* a member of the phylum Mollusca, is a dominant oyster species along the coast waters of South China Sea, with a long cultivation history and has been a popular seafood with a high market demand. The identification of genes related to various physiological processes in this oyster might enhance our understanding of reproduction, possibly contributing to the production of high-quality seeds and also help understand the underlying molecular mechanisms of the traits related to immunity and stress adaptation, which may provide a basis for better strain development in aquaculture.

A total of 147.25 gigabases (Gb) of raw reads were obtained from genome mapping of the Hong Kong oysters by PE125 format of high throughput sequencing on Illumina HiSeq 2500 platform. The estimated final genome assembly (714.88 Mb), covering about 98.20% of the estimated genome size, was found to be composed of 20.34 Kb of contig N50 and 618.24 Kb of scaffold N50 respectively. A total number of 35,624 genes were predicted; of which 90.84% were annotated on the basis of available genomic databases and 1,223 gene families were found to be specific to *C. hongkongensis*. A total of 154 tRNA, 83 rRNA, 807 miRNA, 2,607 pseudogenes and 415.71 Mb repetitive sequences were predicted. Based on comparisons with other all studies of *C. hongkongensis* with other oysters and non-mollusk species it could be concluded that oyster genome might not have undergone large-scale genome duplication events. In addition, the molecular basis of hemolytic phagocytosis and shell formation were also analyzed in genomic level. In conclusion, we report the first draft genome sequence, assembly and annotation of *C. hongkongensis*. The assembled genome will provide a valuable resource for the study of essential physiological processes, phylogeny and evolution among this Hong Kong oyster.

P0284: Aquaculture

Genotyping-by-Sequencing for the Greenshell Mussel Industry

Rachael Ashby[1], Andrew S Hess[1], Hayley Baird[1], Rodney Roberts[2], Nick King[3], Jane Symonds[3], Ken Dodds[1], Neil Gemmell[4] and **Shannon Clarke**[1], (1)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)SPATNZ, Nelson, New Zealand, (3)Cawthron Institute, Nelson, New Zealand, (4)University of Otago, Department of Anatomy, Dunedin, New Zealand

Genotyping by sequencing (GBS) allows aquaculture species with limited availability of genomic tools to establish a low cost genotyping strategy, in turn, enabling the use of genomic selection to drive genetic progress. Furthermore, the genetic diversity can be assessed in already established brood stocks. Here we have used a double digest GBS protocol coupled with construction of genomic relationship matrices (GRM)

to investigate the utility of GBS for parentage and genomic selection for Greenshell™ mussel with the aim of complementing selective breeding programs to optimise traits to deliver benefits for New Zealand's economy.

P0285: Aquaculture

Genetic Diversity and Population Structure in Pacific Lion-Paw Scallop, *Nodipecten subnodosus* from Baja California Peninsula, Mexico.

Cristian Araneda, Universidad de Chile, Depto. de Produccion Animal, Santiago, Chile, Fabiola Lafarga-De La Cruz, Departamento de Acuicultura, CICESE, Ensenada, Mexico, Brenda Oviedo, Postgrado en Ciencias en Acuicultura, CICESE, Ensenada, Mexico, María Eugenia López Dinamarca, University of Chile, Santiago, Chile, Manuel H. Vargas, Postgrado en Ciencias en acuicultura, CICESE, Ensenada, Mexico, María Angélica Larraín, Universidad de Chile, Departamento de Ciencia de los Alimentos y Tecnología Química, Santiago, Chile and Clara E Galindo, Departamento de Biotecnologia Marina, CICESE, Ensenada, Mexico

Genomics technology allows development of thousands of SNP markers in non-model species in a fast and affordable way, opening the opportunity to study population structure in marine mollusk like Pacific lion-paw scallop. We studied 149 scallops sampled in six locations from Baja California Peninsula in Mexico (five from Gulf of California coasts and one from Pacific coast of California) along with progenies from one population obtained from a Gulf of California (aquaculture production) and genotyped with 1,190 SNP loci obtained by RAD-seq. Genetic diversity was low and similar across locations ($0.065 < He < 0.075$). Differentiation among locations using 627 putatively neutral loci was very low as expected ($F_{ST} = 0.0032$) and it was very similar to the obtained with 62 putative outlier loci ($F_{ST} = 0.0042$), indicating a panmictic population in this area. ADMIXTURE analysis with the full SNPs panel (1,190 loci), 62 outlier loci and 627 neutral loci showed a similar pattern with two ancestries shared among locations (k =2), but one of them with high representation in samples (more than 75%). Interestingly, cultured Pacific lion-paw scallop show an inverse pattern of ancestry, putting in evidence the effect of aquaculture practices on genetic diversity from fisheries resources.

P0286: Aquaculture

The Red Abalone Genome and the Consequences of Inbreeding in White Abalone Aquaculture

Rick Masonbrink[1], Arun S. Seetharam[1], Sivanandan Chudalayandi[1], John Hyde[2], Catherine Purcell[2] and Andrew J. Severin[1], (1)Iowa State University, Ames, IA, (2)Southwest Fisheries Science Center, National Marine Fisheries Service, La Jolla, CA

Abalone aquaculture production is rapidly growing ~40% annually (40k tons in 2008), propelled by both increases in demand and the devastation of wild populations. While abalone aquaculture continues to grow, progress is limited by the lack of sex markers and bottlenecks limiting growth, and production efficiency. Genomic resources have been leveraged to identify traits important for abalone production, environmental tolerance, disease resistance, and sex determination. As an initial step toward this goal a red abalone *(Haliotis rufescens*) genome assembly was generated using MaSuRCA (v 3.22) with both Illumina (80X) and PacBio reads (20x), and further scaffolded using Hi-C data with Dovetail. The assembly is comprised of 8405 scaffolds with a ~1.9Mb N50, and a longest scaffold of 13.2 Mb. The assembled genome (1.5Gb) size is smaller than the predicted size (1.8Gb), yet genome completeness was high at 95.3% complete with BUSCO. These stats are further bolstered by raw read alignments accounting for 91% and 86% for PacBio and Chicago reads, respectively. A GWAS study has been implemented to identify sex-determination loci, using male and female samples from 6 abalone species. Additionally, to identify bottlenecks in cultured populations, GWAS was utilized with 11 wild and 11 cultured white abalone (*H. sorenseni*) samples.

P0287: Microbes and Pathogens

Draft Genome Sequence of the Pearl Millet Downy Mildew Pathogen

Rakesh K. Srivastava, ICRISAT, Greater Hyderabad, India

Sclerospora graminicola pathogen causing downy mildew disease is the most important biotic production constraints of pearl millet in India, Africa and other parts of the world. We report a de novo whole genome assembly and analysis of pathotype 1, one of the most virulent pathotypes of *S. graminicola* from India. The whole genome sequencing was performed by sequencing of 7.38 Gb with 73,889,924 paired end reads from the paired-end library, and 1.15 Gb with 3,851,788 reads from the mate pair library generated from Illumina HiSeq 2500 and Illumina MiSeq, respectively. A total 597,293 filtered sub reads with average read length of 6.39 Kb was generated on PACBIO RSII with P6-C4 chemistry. Assembled draft genome sequence of *S. graminicola* pathotype 1 was 299,901,251 bp in length, N50 of 17,909 bp with a minimum of 1 Kb scaffold size. The GC content was 47.2 % consisting of 26,786 scaffolds with longest scaffold size of 238,843 bp. A total of 52,285 predicted genes found homology using BLASTX against nr database and 38,120 genes were observed with a significant BLASTX match. Out of 38,120 genes annotated a set of 11,873 genes had UniProt entries, while 7,248 were GO terms and 9,686 with KEGG IDs. Of the 7,248 GO terms, 2,724 were associated with the biological processes. The Sclerospora graminicola whole genome shotgun (WGS) project is available under BioProject ID PRJNA325098. This study may help understand the evolutionary pattern of pathogen and aid elucidation of effector evolution for devising effective durable resistance breeding strategies in pearl millet.

P0288: Microbes and Pathogens

Reconstruction of Metagenome-Assembled Microbial Genomes from a Microbiome

Bhavya Papudeshi, Indiana University, Bloomington, IN and Elizabeth Dinsdale, San Diego State University, San Diego, CA

Microbiome/host interactions describe characteristics that affect the host health; shotgun metagenomics sequences microbiome samples, allowing us to analyze its taxonomic and metabolic potential. Reconstruction of metagenome fragments into genomes (called metagenome-assembled genomes) that facilitates linking function to taxa within microbial symbionts. Reconstruction of genomes sort assembled sequences into bins, characteristic of a genome. However, the microbial community composition, including taxonomic and phylogenetic diversity may influence genome reconstruction. We determine the optimal reconstruction method for four microbiome projects with variable sequencing platforms, diversity, and environment using a set of parameters to select for optimal assembly and binning tools. We evaluated 3 assemblers (IDBA, MetaVelvet, and SPAdes) and 2 binning tools (GroopM and MetaBat) for four projects (105 metagenomes). We find that SPAdes

assembled more contigs (143,718 ± 124) of longer length (N50 = 1632 ± 108 bp), incorporated the most sequences (19.65 %), and low chimera levels (microbial richness and evenness were maintained across assembly). SPAdes assembly was responsive to biological and technological variations within the projects. MetaBat binning tool produced bins, characteristic of a genome with less GC variation (standard deviation 1.49), low species richness (4.91 ± 0.66), and higher genome completeness (40.92 ± 1.75). MetaBat extracted 115 bins of which 66 bins were identified as quality reconstructed metagenome-assembled genomes with a genus specific sequences. In conclusion, we present a set of biologically relevant parameters to select for optimal assembly and binning tools. SPAdes and MetaBat tools reconstructed quality metagenome-assembled genomes for the four projects included in this study.

P0289: Microbes and Pathogens

SMART: Simple Microbiome Analysis Research Toolkit – Online Discovery and Visualization Workflows for Non-Informaticians Using Existing Tools

Kristopher Irizarry, The Applied Genomics Center, College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA and Brian Oakley, Microbial Ecology, College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA

Simple **M**icrobiome **A**nalysis **R**esearch **T**oolkit (SMART) is a modular, web-based, bioinformatics pipeline for rapidly converting microbial 16S rRNA sequence data into visually rich knowledge maps encoding functionally relevant signals within the underlying metagenome. SMART is unique in that it instantly enables non-bioinformaticians to harness all of the features of highly-customizable bioinformatics discovery engines without requiring users to have computational expertise or access to linux/unix computers. In fact SMART was specifically developed for doctors, researchers, students, veterinarians and biotechnology, pharmaceutical and agricultural professionals. Because many tools present a barrier to entry for non-computational users, this toolkit was specifically developed to provide a user-friendly, hypothesis-driven, genomics discovery pipeline from a set of existing publically accessible, and previously described web-based informatics resources. Furthermore, the algorithmic components of the SMART pipeline and associated visualization engines were intentionally selected from among existing online resources to capitalize on the processing capabilities of the human visual system which simultaneously categorize data-rich sensory streams of visual information into spatial patterns and feature sets. SMART presents visual data in ways that are most easily perceived and digested by the human user. Ultimately, this leads to enhanced quality and quantity of information exchange at the human-software interface which translates directly to enhanced extraction of biologically relevant signals from microbiome data sets.

P0290: Microbes and Pathogens

Understanding the Mechanism of *E. coli* Resistance to Membrane-Targeting Antimicrobial through Genomic Approach

Roel C. Rabara[1], Loan Huynh[1], Jeanette Velasquez[1], Hau Nguyen[2], Supratim Basu[1], Shujian Zhang[1] and Goutam Gupta[1], (1)New Mexico Consortium, Los Alamos, NM, (2)NEW MEXICO CONSORTIUM, LOS ALAMOS, NM

Antimicrobial peptides (AMPs) are essential component of barrier defense system in plants that kill bacteria through interactions with phospholipids and membrane permeabilization. However, bacteria develop resistance against AMPs by gaining mutations that help overcome interactions with phospholipids and membrane permeabilization. To determine the nature of bacterial resistance against AMPs, we used *Escherichia coli* strain BL21-Gold (DE3) as a model bacterium to evolve resistance against an amphipathic 11 amino acid long helical peptide (or P11). The minimal inhibitory concentration (MIC) of the resistant strain is 13-fold higher than the wildtype (or susceptible) strain. Genome sequencing of the resistant *E. coli* strain revealed several insertions and deletions in multiple genes. Through genome analysis, we observed a transposase insertion in the genes that included *asmA*, *dusC* and *waaP/rfaP*. The gene *asmA* encodes the assembly protein asmA that is involved outer membrane assembly while *waaP/rfaP* encodes the enzyme lipopolysaccharide core heptose (I) kinase that is required in the addition of phosphate to O-4 of the heptose I residue in lipopolysaccharide (LPS) core. *Dihydrouridine synthase C* (*dusC*) encodes the enzyme tRNA-dihydrouridine[16] synthase that is involved in 5,6-dihydrouridine modification observed in wild-type cellular tRNA specifically the modification of U16 of the D loop in tRNAs. Knocking out these three genes resulted in 3-fold increase of the MIC compared to the control. In addition, there are mutations in other genes that encode proteins involved in interactions with phospholipids and membrane permeabilization of P11. All the mutated genes account for the total resistance. Finally, we have been able to design the next-genera AMP, a 26 amino acid long peptide (or P26) that is not susceptible to the resistance mechanism.

P0291: Microbes and Pathogens

Single Transposon-Mediated Gene Capture Event Preceded Horizontal Transfer of Virulence between Multiple Fungal Pathogens

Adam P. Taranto[1], Megan C. McDonald[1], Andrew W. Milgate[2] and Peter S. Solomon[1], (1)The Australian National University, Canberra, Australia, (2)NSW Dept of Primary Industries, Wagga Wagga, Australia

Fungal pathogens deploy effector proteins to manipulate host immune responses and facilitate infection. Effectors frequently show discontinuous distributions between species suggestive of horizontal gene transfer (HGT) between species sharing a common host. The effector *ToxA* of the wheat pathogen *Parastagonospora nodorum* is a classic example, conveying virulence upon non-pathogenic isolates and notably contributing to the emergence of a new wheat pathogen, *Pyrenophora tritici-repentis* by means of HGT. Effector genes, such as *ToxA,* are commonly found in association with transposable elements (TEs) and it has been hypothesised that such elements are a factor contributing to interspecific transfer. However, no specific TE has yet been identified as the unit of transfer in a case of horizontal effector transmission. We have recently discovered *ToxA* in wheat-infecting isolates of a third species, *Bipolaris sorokiniana*. Chromosome-level assemblies for multiple *B. sorokiniana* and *P. nodorum* isolates enabled us to identify extended regions of shared identity around *ToxA*. We have subsequently reconstructed a single composite DNA-transposon responsible for mobilising *ToxA* between species and show this element to be intact and active in *B. sorokiniana* - this is the first time a functional TE has been definitively linked to a case of HGT in fungi. In each species, *ToxA* resides in a large TE-rich genomic island (aka AT-isochore or LS region). We propose that rare inter-specific hybridisation events facilitate introgression of mobile-element islands into new populations and that individual elements act as rafts mobilising captured effectors between islands.

P0292: Microbes and Pathogens

Comparative Genome and RNA-Seq Analysis of *Colletotrichum scovillei* (*C. acutatum* sensu lato)

Kyoung Su Kim, Kangwon National University, Chuncheon, Korea, Republic of (South)

An ascomycete plant pathogen, *Colletotrichum scovillei* (*C. acutatum* sensu lato) is a major threat in pepper production worldwide. Genome-wide comparative analysis of *C. scovillei* with species in *Colletotrichum* and outgroup fungi has been performed. Based on comparative analysis, we extended in silico detailed analysis of TFome and secretome of S. scovillei, which classified 601 and 1323 genes, respectively. In addition, 464 genes were identified to encode small secreted proteins, among which 39 genes was *C. scovillei*-specific proteins, compared to compared species of *Colletotrichum*. Functional analysis of several genes in secretome using knockout mutants and fluorescent microscopic examination will be presented. Genes encoding homeobox domain were targeted to generate mutants. Functional roles of those genes will be included together with RNA seq data.

P0293: Microbes and Pathogens

Rapid DNA Isolation from Diverse Plant Material for use in Next Generation Sequencing Applications

Heather Martinez, Nicola Scholle, Markus Sprenger-Haussels and Dominic O'Neil, QIAGEN GmbH, Hilden, Germany

Isolation of DNA from plant material is often a tedious process which involves significant hands on time and leads to varying results due to the diverse nature of the material. Different parts of the plants as well as the plants themselves differ in both consistency of material and presence of inhibitory substances, making dependable isolation of DNA difficult

Here, we developed a method for the efficient extraction of DNA from different plant types, including strawberry leaf, pine needles, grape leaf, and cotton and coffee seeds. More efficient sample lysis with minimal hands-on time leads to significantly increased DNA yield compared to conventional methods. Through the use of multiple technologies to improve removal of small molecules, such as polyphenols and complex polysaccharides that may inhibit downstream applications, the isolated DNA is of high quality and purity.

The resulting DNA is suitable for immediate use in downstream reactions, including PCR, qPCR and Next Generation Sequencing based applications. Using this method we were further able to design a workflow that included DNA isolation, library preparation and bioinformatics analyses for the efficient detection of plant pathogens isolated from infected samples. With this, our protocol is an substantial improvement within workflows used for plant microbiome and plant pathology studies as well as in plant breeding and engineering.

P0294: Microbes and Pathogens

Measuring Comparative Performance of the Applied Biosystems™ Axiom™ Microbiome Array with Respect to 16S rRNA and Shotgun Metagenomic Sequencing for a UK Biobank Pilot Cohort

Ali Pirani[1], Manasi Shah[1], Joseph Foster[1], Olga Kamneva[1], Laura Cobb[2], Samantha Welsh[2], Graham Tilston[2], Clair Bloor[1], Geoff Scopes[1] and Rachael Almond[2], (1)Thermo Fisher Scientific, Santa Clara, CA, (2)UK Biobank, Adswood, United Kingdom

Microbial community detection in medical, agricultural, and ecological specimens is crucial for addressing infectious diseases, crop yield, pathogens, and bioremediation, among other natural phenomena.

Targeted next generation sequencing (NGS) of 16S rRNA gene markers is affected by read length, sequencing depth and reference database for annotation, which limits finer resolution taxonomic assignment. On the other hand, whole genome shotgun sequencing is expensive and requires substantial bioinformatics expertise.

The Applied Biosystems™ Axiom™ Microbiome Array (Axiom) provides an attractive alternative platform detecting over 12,000 organisms including archaea, bacteria (species and strains), fungi, protozoa, and viruses with probes targeting whole genome sequences.

The array content is sample type agnostic and data can be analyzed with easy-to-use Applied Biosystems™ Axiom™ Microbial Detection Analysis Software (MiDAS).

As part of a UK Biobank potential sample collection protocol, a pilot study was conducted to determine optimal sample collection, storage, and treatment methods for microbial community estimation. We performed a unique comparison of three technologies: Axiom™ Microbiome array, targeted 16S rRNA NGS and whole genome sequencing (WGS) for characterizing microbial composition in this study.

P0295: Microbes and Pathogens

Genome Analysis of the Endophytic Fungi Producing the Anticancer Drug Taxol

Jinhee Choi and **Kwang-Hyun Baek**, Department of Biotechnology, Yeungnam University, Gyeongsan, South Korea

Taxol has been used as the most extensively used chemotherapeutic drug for treatment of many cancers. After being isolated from *Taxus* sp., many endophytic fungi have been known to produce taxol. *Taxomyces andreanae* was the first isolated fungi producing taxol in 1993, and many endophytic fungi were isolated from various plant species. The amount of taxol from the endophytic fungi cannot be used for the commercial production of taxol, however, the simple genome structure and easy gene manipulation of the fungi is the most considerable factors for identifying the unknown genes in the taxol biosynthetic pathway. We have collected many endophytic species as taxol accumulators, and successfully sequenced the genomes of *Colletotrichum dematium* and *Cladosporium cladosporioides*. We anticipate that the functional genome analysis of the taxol-producing endophytic fungi can elucidate the gene clusters for the taxol biosynthetic pathway.

P0296: Microbes and Pathogens

Novel Similarities in Transcriptional Gene Expression Regulation during Basal Immunity in Mammals and Plants

Konstantin Kanofsky, Alexander Sandmann, Anika Möller and Reinhard Hehl, Institut für Genetik, Technische Universität Braunschweig, Braunschweig, Germany

One of the basic questions in innate immunity in animals and plants concerns its origin. Did innate immunity originate in an ancient unicellular organism by divergent or convergent evolution? In both kingdoms recognition of conserved molecules associated with pathogens leads to basal immunity accompanied by the upregulation of numerous pathogen response genes. In contrast the expression of pathogen defense genes depends on different transcription factors in mammals and plants. No similarities in transcriptional regulation in animals and plants have been

found so far. A novel *A. thaliana* microbe-associated molecular pattern (MAMP)-responsive *cis*-sequence, designated type II WT-box, is required for flg22-responsive gene expression in protoplasts. Surprisingly, a bioinformatic analysis to identify putative binding proteins predicts mouse (*Mus musculus*) NF-κB p65 as a binding transcription factor. NF-κB p65 plays a role in innate immunity in mammals and a homologue in *A. thaliana* is not known so far. The occurrence of NF-κB p65 target sites in MAMP-responsive sequences in *A. thaliana* was astonishing. Therefore, the interaction of NF-κB p65 with these target sites was tested experimentally. NF-κB p65 activates WT-box containing synthetic promoters in plant cells. Activation requires the WT-box and NF-κB p65 binds to WT-box containing synthetic promoters *in vitro*. These results uncover a further similarity in innate immunity between mammals and plants, in this case at the transcriptional level.

P0297: Microbes and Pathogens

Host Genetic Background Drives Developmental Trajectories in Hosts with Disrupted Microbiota

Kathryn CA Milligan-Myhre and Emily Lescak, University of Alaska Anchorage, Anchorage, AK

Resident microbiota play an important role in host health, and have complex relationships with their host. Host microbe interactions can be influenced by several factors, including the immune response to microbes and host genetic background. To determine what role the host genetic background plays in host-microbe interactions, we adapted well developed evolution animal model, threespine stickleback (*Gasterosteus aculeatus)* as a model host for host-microbe interaction studies. Threespine stickleback are well characterized, have an annotate sequenced genome, and, important for our studies, have undergone parallel evolution as ancestral oceanic populations have adapted to freshwater environments around the northern hemisphere. In Alaska we have ready access, and sample, populations of stickleback that have evolved in hundreds of different microbial, abiotic, and macrobiotic environments. We hypothesize that these populations have evolved different relationships with their microbiota. To test this hypothesis, we have examined the immune response and somatic development of several populations raised in the presence of complex microbial communities, absence of microbes, or with a microbiota challenged with antibiotics. We have found that stickleback that have evolved in different environments indeed have different immune responses and developmental trajectories when their microbiota is disrupted. We are now expanding our studies to new populations, developing new assays to measure developmental changes, and determining the role of stress in this delicate balance between the host and its microbial community.

P0298: Microbes and Pathogens

Rediscovering Broad-Spectrum Disease Resistance in Rice: Prompting a Genome-Wide Uprising

Bradley Tonnessen[1], Ramil Mauleon[2], Nickolai Alexandrov[2] and Jan E. Leach[3], (1)Colorado State University, Fair Oaks, CA, (2)International Rice Research Institute, Los Baños, Philippines, (3)Colorado State University, Fort Collins, CO

Rice is the main staple food crop of the world, and thus, losses caused by diseases threaten international food security. Single resistance genes offer effective disease control, but are inadequate due to rapid pathogen evolution. There is a need for durable resistance, which is frequently quantitative (multigenic), and involves Defense Response (DR) genes that are effective against diverse pathogens. In this work, we sought to understand how expression of the many DR genes involved in resistance is orchestrated, with the long term goal of enabling genome-wide breeding for more effective resistance. We identified DR gene promoter short sequence motifs that are shared across co-expressed, broad spectrum DR genes. Specific groupings of these DR-associated motifs, known as *Cis*-Regulatory Modules (CRMs), are present in many promoters, and are enriched in DR genes. These CRMs harbor *cis*-elements known to be involved in disease resistance, and some are involved in small RNA expression, putatively contributing to epigenetic regulation of many DR genes. In known disease resistance Quantitative Trait Loci (QTL), polymorhpisms in CRMs occur in resistant relative to susceptible QTL haplotypes in DR gene promoters. The presence of intact CRMs exclusively in resistant haplotypes of DR genes provides evidence that these CRMs have a predicative role in the contribution of other DR genes to resistance. Based on this information, we predict that a CRM signature within DR gene promoters can be used as a marker for future breeding practices to enrich for the most responsive and effective DR genes across the genome.

P0299: Microbes and Pathogens

Assessing the Potential of Host-Induced Gene Silencing to Reduce Wheat Rust Infection in Transgenic Wheat

Ali A. Ahmed, UC DAVIS, DAVIS, CA and Chuntao Yin2, Wenjun Zhang1, Jorge Dubcovsky1, Scot Hulbert2, Richard Michelmore1 1 UC Davis, Davis, CA, USA. 2 Washington State University, Pullman, WA, USA

The RNA interference based Host-Induced Gene Silencing (HIGS) has been shown to provide resistance to several plant diseases. We are testing HIGS to enhance resistance in wheat against three rust pathogens. Previously we demonstrated that individual silencing of ten stem rust genes using transient Virus-Induced Gene Silencing significantly reduced stem rust development. Silencing four and three of these genes also decreased the development of stripe rust and leaf rust, respectively. We are currently investigating the potential of silencing these genes to minimize rust infection through stable expression of the RNAi trigger sequence in transgenic wheat. Four stripe rust genes which are highly expressed in haustoria and two leaf rust genes that may be vital for proliferation were also targeted. So far 154 T1 plants have been generated with an average of six independent transgenic events per target gene, of which 140 have been confirmed as containing the respective transgene by PCR. The T1 plants were selfed to generate T2 seed. Ten T2 plants per each T1 are being evaluated for resistance to stripe rust and stem rust on 1-9 and 0-4 scales, respectively. Of 68 T1 progeny tested so far against stripe rust, 6 and 7 infection types were observed for 3 and 13 transgenics, respectively. One, nine and four transgenics showed 2-3, 3 and mixed infection types for stem rust infection, respectively. Seven plants showed some level of reduced infection with both stripe rust and stem rust. The infection assays are underway for the remaining T1 plants. Tests will be repeated to confirm the partial resistance phenotypes on T3 plants. All 140 T1 progeny will also be evaluated for resistance to leaf rust. Transgenic lines containing different RNAi targets that exhibit promising phenotypes will be crossed in order to pyramid individual RNAi effects.

P0300: Microbes and Pathogens

Atypical Infection By *Moniliophthora perniciosa* Reprograms Primary Metabolism and Generates Sink in Symptomatic Tissues of "Micro-Tom"

Daniele Paschoal[1], Juliana Leles Costa[1], Paulo José Pereira Lima Teixeira[2], Camila Caldana[3], Valdemar Tornisielo[4], Juliana Aparecida Aricetti[5], Rodrigo Floriano Pimpinato[4] and Antonio Figueira[6], (1)Plant Breeding Laboratory, Center for Nuclear Energy in Agriculture, University of Sao Paulo, Piracicaba, Brazil, (2)University of North Carolina at Chapel Hill, Chapel Hill, NC, (3)Max Planck Institute for Molecular Plant Physiology, Potsdam, Germany, (4)Ecotoxicology Laboratory, Center for Nuclear Energy in Agriculture, University of Sao Paulo, Piracicaba, Brazil, (5)Brazilian Bioethanol Science and Technology Laboratory, Campinas, Brazil, (6)University of Sao Paulo, Piracicaba, Sao Paulo, Brazil
Witches' broom disease is a devastating cacao disease in South America. The causal agent is the hemibiotrophic basidiomycete *Moniliophthora perniciosa* with a peculiar long biotrophic phase. The availability of an S-biotype that colonizes tomato enabled the use of 'Micro-Tom' (MT) as a model system to investigate the interaction. Hypertrophic growth of stems and proliferation of axillary shoots are typical symptoms of cacao and MT infection, which suggest a hormonal and metabolic imbalance. Therefore, we analyzed infected MT by RNA-seq for expression of genes associated with primary metabolism. Symptomatic tissues showed a metabolic reprogramming with induction of genes encoding hexose-phosphorylases and transporters, sugar sensing and signaling enzymes (such as trehalose-6-phosphate), and repression of those associated with photosynthesis (Calvin cycle) and starch biosynthesis. TOF-GC-MS analysis revealed the accumulation of organic acids from the Tricarboxylic Acid Cycle, corresponding to an increase in respiration, whereas the accumulation of orthophosphate, polyamines, sugar acids, aldohexoses and fructose, indicated a highly metabolic active tissue. Major translocation of ^{14}C-glucose applied to leaves to symptomatic regions indicated the occurrence of a metabolic sink. Application of benzyl-adenine to shoot apices induced symptoms similar to infection, with the corresponding increase in translocation of ^{14}C-glucose, suggesting a role for cytokinin in sugar translocation. Further, infection of 35S::AtCKX2, a transgenic line with low cytokinin by *M. perniciosa* did not direct translocation of ^{14}C-glucose to infected stems. Our results indicate a shift in primary metabolism during infection by M. *perniciosa*, with a presumed role of cytokinin on inducing metabolite sink towards the symptomatic region.

P0301: Microbes and Pathogens

Transcription Profiling of the Infection Process of *Botrytis cinerea* on Whole Tomato Plant Leaves

Dhruv Aditya Srivastava[1], Eswari Pandaranayaka PJ[1], Omer Frenkel[2], Dov Prusky[3], Yigal Elad[2] and **Arye Harel**[1], (1)Vegetable and Field Crops ,Institute of Plant Sciences, Agricultural Research Organization (ARO), Volcani, Rishon Lezion, Israel, (2)Department of Plant Pathology and Weed Research, Agricultural Research Organization (ARO), Volcani, Rishon Lezion, Israel, (3)The Department of Postharvest and Food Sciences, Agricultural Research Organization (ARO), Volcani, Rishon Lezion, Israel
Botrytis cinerea is a foliar necrotrophic fungal-pathogen which is capable of infecting over 1,400 plant species [1], and was ranked second worldwide for its scientific and economic importance [2]. In spite of the importance of this pathogen, a transcription profiling of its infection process on a whole plant (as opposed to detached tissue) of widely used crop (except for cucumber and lettuce) was not studied. We analyzed the transcriptome of *B. cinerea* (strain B05.10) infection on one of the most important vegetable crops in the world, tomato (*Solanum lycopersicum*, cv. M82). We sampled six week old infected leaf tissues at 0, 16, 23, 40, and 48 hours post infection. Approximately 35, or 45% of *B. cineria* or *S. lycopersicum* genes were differentially expressed during pathogenicity, respectively, demonstrating the global effect of this process. Preliminary KEGG enrichment analysis of *B. cinerea* transcriptome illustrated over-expression of genes involved in regulation of transcription, translation, DNA repair and recombination, in early stages of the infection. While genes involved in interaction with the environment, and plant secondary metabolite synthesis pathways (e.g., phenylpropanoid and iIsoquinoline alkaloid biosynthes) were upregulated in the later part of the infection (i.e., establishment). The latter could illustrate how *B. cinerea* manipulates plant growth/defense systems to accomplish establishment. Altogether, our analysis and its future validation may increase our understanding on plant-fungal interactions which are essential for successful infection.

[1] Elad, Y., Pertot, I., Cotes Prado, A.M., and Stewart, A. (2016) Plant hosts of *Botrytis* spp. In *Botrytis - the Fungus, the Pathogen and its Management in Agricultural Systems*.

[2] Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D. et al. (2012) The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol* **13**: 414-430.

Equal contribution - D.A. Srivastava and E. Pandaranayaka PJ

P0302: Microbes and Pathogens

Comparative Transcriptome Analysis of Susceptible and Resistant Tomato (*Solanum lycopersicum*) Genotypes from Ghana to the Root-Knot Nematode

Seloame T. Nyaku[1], Uta Paszkowski[2] and Krys Kelly[2], (1)University of Ghana, Accra, Ghana, (2)University of Cambridge, United Kingdom
Tomato (*Solanum lycopersicum* L.) is utilized in many Ghanaian dishes. However, its yields are limited by the nematode *Meloidogyne incognita*. There is therefore the need for novel control measures, which require molecular mechanisms underlying the plant defense responses to *M. incognita* to be thoroughly understood. Our objectives were to analyze site-specific rhizogenic gene expression in four *Solanum* species ('MongalT-11', 'Pectomech', 'VFN8' and *S. galapangense*) against *M. incognita* infection during both compatible and incompatible reactions, and to characterize the response mechanisms and plant defense signaling pathways during infection. A split-root technique involving half of the root system inoculated with juveniles of *M. incognita* was utilized, and roots harvested 3, 5, 12, 15, and 21 days after inoculation (dai). The 18S (nematode-specific primer, UI416988(Jasmonic-specific primer), ACS2 (Ethylene-specific primer), PR1 (Pathogenesis-related gene-1 for salicylic acid) and Ubiquitin-Ubi3 (House-keeping gene) were used in quantitative RTPCR to comparatively determine the expression of ethylene, salicylic acid, and jasmonic acid-regulated marker genes. Sequencing of cDNA libraries for these genotypes were achieved using Illumina NextSeq® 500 platform, which will enhance our understanding of *M. incognita* interactions in the context of expressed genes and signaling pathways, and comparative defense responses among the tomato species.

P0303: Microbes and Pathogens

Genome Sequencing of *Thecaphora frezii* (Peanut Smut) and Comparative Genomics to *Ustilago maydis*

Renee S. Arias[1], Alicia N Massa[2], Luis I. Cazon[3], Alejandro M. Rago[4], Mary V. Duke[5], Linda L. Ballard[6], Sheron A. Simpson[7], Victor S. Sobolev[1], Juan A. Paredes[8], Mario I. Buteler[9] and Brian E. Scheffler[5], (1)USDA-ARS National Peanut Research Laboratory, Dawson, GA, (2)USDA-ARS-National Peanut Research Laboratory, Dawson, GA, (3)Instituto Nacional de Tecnologia Agropecuaria (IPAVE), Cordoba, Argentina, (4)IPAVE-CIAP-INTA, Cordoba, Argentina, (5)USDA-ARS, Stoneville, MS, (6)USDA-ARS-SEA, Stoneville, MS, (7)USDA-ARS, Genomics and Bioinformatics Research Unit, Stoneville, MS, (8)INTA-IPAVE, Cordoba, Argentina, (9)Criadero El Carmen, Cordoba, Argentina

Peanut "smut" disease can cause up to 50% losses of this crop in the field. Peanut smut is caused by the fungal pathogen *Thecaphora frezii*. Since knowledge about the genetics of plant pathogens helps to develop management strategies, we initiated a project for the whole genome sequencing of this organism. Currently, in NCBI (National Center for Biotechnology Information) only has 10 nucleotide entries for this species for a total of 6500 bp of ribosomal RNA and ITS (internal transcribed spacer). In a combined effort between USDA-ARS-National Peanut Research Laboratory and INTA (Instituto Nacional de Tecnologia Agropecuaria, Argentina), we sequenced the genome of *Thecaphora frezii* using a combination of sequencing platforms, Pacific Biosystems (PacBio) and Illumina. PacBio rendered 980894 reads (average length of 7,388 bases) with a total of 1382 Mb, and Illumina resulted in 75.8 M paired end reads (average length of 188 bases) with a total of 1413 Mb. De novo assembly and mapping to a known genome of a related species*, Ustilago maydis*, were performed. Comparisons of regions potentially related to pathogenicity are discussed.

P0304: Microbes and Pathogens

Community Assortment of Microbes Associated with Wild Chickpea

Betsy Alford[1], Alex Greenspan[1], Eric Bishop-von Wettberg[2], R. Varma Penmetsa[1] and Douglas R Cook[1], (1)University of California-Davis, Davis, CA, (2)University of Vermont, Burlington, VT

Most food crops are cultivated in areas distant from their center of origin. However, it is uncertain whether microbes from the center of origin were “transplanted” as crops were moved to new areas, or whether and to what extent native microbes in new areas of cultivation yield surrogate functionalities. Additionally, whether microbial communities were co-domesticated with crops and selected for crop-relevant adaptations is unknown. Answering such questions requires that we first understand the plant microbiome in native communities of crop wild progenitors.

To build a comprehensive model of the wild chickpea microbial communities, microbial DNA, soil and plant tissue was collected in increasing proximity to the individual plants. To determine the effects of planting site, twenty sites were sampled over the geographic range of both wild species *C. reticulatum* and *C. echinospermum*. Abiotic information was collected such as soil chemistry and weather data with in-soil sensors. Microbial communities demonstrated a selection gradient moving from the soil toward the plant compartments. Moving from soil to root, there is a marked increase in Acidiobacteria and Actinobacteria, additionally there is a decrease in percent of Proteobacteria. Community analysis of the wild system demonstrated that plant genotype, soil type and symbiont colonization co vary with microbial community assortment, however, a follow up common garden experiment using wild plants and soils revealed that community assortment is driven by soil type.

P0305: Microbes and Pathogens

Global Population Genomics of Chickpea-Nodulating Mesorhizobium

Alex Greenspan[1], Betsy Alford[1], Peter Chang[2], Eric Bishop-von Wettberg[3], Abdullah Kahraman[4], Asnake Fikre[5] and Douglas R Cook[1], (1)University of California-Davis, Davis, CA, (2)University of Southern California, Los Angeles, CA, (3)University of Vermont, Burlington, VT, (4)Harran University, Urfa, AZ, Turkey, (5)Ethiopian Institute of Agricultural Research, Addis Ababa, Ethiopia

Legume crops are significant agriculturally and environmentally for their ability to form symbiosis with specific soil bacteria capable of nitrogen fixation. Nitrogen fixation for a given legume in a given soil is limited by the availability of the plant’s bacterial partners, and by variation in the effectiveness of those symbionts. We used a global-level hierarchical sampling scheme to comprehensively characterize the evolutionary relationships and distributional limitations of nitrogen-fixing bacterial symbionts of the legume crop chickpea. This has been accomplished using culture-dependent and independent approaches to generate over 1,200 draft whole-genome assemblies at the level of bacterial populations, as well as 17 finished-quality genomes using the Pacific Biosciences platform. These strategies reveal that that chickpea’s symbionts across the globe are confined to the genus *Mesorhizobium*, but a diversity of taxa within the genus. Comparative phylogenomic analysis reveals that despite chickpea’s symbionts within and across regions coming from different taxa, all share almost identical genes for symbiosis. PacBio genome-assemblies reveal that this is due to the horizontal transfer of a 500 kb chromosomal island known as a symbiosis island, between unrelated strains of the genus *Mesorhizobium*. Analyzing the symbiosis island at the population level reveals that the symbiosis island spreads repeatedly once introduced to a region, suggesting that strains well-adapted to a particular soil climate continue to dominate once the new host (chickpea) has been introduced, through repeated acquisition of the symbiosis island. This dataset provides additional insights into the functional and taxonomic diversity of other bacteria associated with chickpea nodules.

P0306: Microbes and Pathogens

Candidate Genes for Resistance to *Fusarium oxysporum* f.sp. *cubense* race 1 in the Wild Diploid Banana Species *Musa acuminata* var. *malaccensis*

Fajarudin Ahmad[1,2], Yuyu S Poerba[2], Hugo A Volkaert[3], Harold JG Meijer[1], Hans de Jong[1], Henk Schouten[1] and Gert HJ Kema[1], (1)Wageningen University & Research, Wageningen, Netherlands, (2)Research Center for Biology, Indonesian Institute of Sciences, Bogor, Indonesia, (3)Center for Agricultural Biotechnology, Kasetsart University, Kamphaengsaen, Thailand

Banana is one of the most important fruit crops in the world and a staple food in many tropical countries. The current annual production is over 100 million tons, but this is threatened by one of the most devastating fungal diseases: Fusarium wilt. The causal agent is *Fusarium oxysporum* f.sp. *cubense* (*Foc*). During the 1950s the “Gros Michel” banana based industry was devastated by *Foc* Race 1. However, the epidemic was quenched and the banana industry was saved by replacing “Gros Michel” with resistant “Cavendish” bananas. Surprisingly, the responsible

gene(s) for resistance are still unknown. We used the self-compatible wild banana accession *Musa acuminata* var. *malaccensis* (*Mam,* AA, 2n=22) from the Sumatra *Musa* population to generate a mapping population and to investigate the inheritance of resistance to *Foc* Race 1. Initial greenhouse bio-assays confirmed that *Mam* is resistant to Race 1. The F1 population was generated from 272 pollinated flowers and resulted in 3,458 seeds of which 718 were embryo rescued, and eventually 255 off-spring survived and were maintained in tissue culture. The mapping population was genotyped (N=244) using DArTseq markers and subsequent phenotyping (N=225) revealed segregation for resistance. After strict filtering, 4,171 SNP markers were used for genetic mapping. Analyses of the genotyping and phenotyping data showed the inheritance of a single dominant resistance gene that mapped near the distal part of chromosome 10, based on the reference genome of doubled haploid 'Pahang', which is also a *Mam* accession. The recombination between the markers among the selected recombinants, together with the position of the putative resistance gene, were further analyzed using graphical genotyping and resulted in markers that flank a 360 kb genetic region containing at least 14 NBS-LRR like resistance gene candidates, including the identified candidate gene for resistance to *Foc* Race 1.

P0308: Microbes and Pathogens

Development of Diagnostics for Yam Virus Detection

Chukwuemeka K. Nkere[1], Gabriel I. Atiri[2], Joseph Onyeka[3], Chiedozie N. Egesi[3], P. Lava Kumar[4] and Susan E. Seal[5], (1)National Root Crops Research Institute (NRCRI) Umudike, Umuahia, Nigeria, (2)University of Ibadan, Ibadan, Nigeria, (3)National Root Crops Research Institute (NRCRI) Umudike, Umuahia, Nigeria, (4)International Institute of Tropical Agriculture, Ibadan, Nigeria, (5)NRI, University of Greenwich, Chatham Maritime, United Kingdom

Yam (*Dioscorea* spp.) is a vegetatively propagated crop and important food staple in West Africa which is susceptible to pests and diseases, particularly tuber-borne viruses. *Yam mosaic virus* (YMV, genus *Potyvirus*), *Yam mild mosaic virus* (YMMV, genus *Potyvirus*), *Cucumber mosaic virus* (genus, *Cucumovirus*) and *Dioscorea bacilliform virus* (DBV, genus *Badnavirus*) are the major viruses infecting yams in West Africa with high incidence, prompting the need for an efficient diagnostic tool for strain and species discrimination in these genera. We developed a multiplex RT-PCR assay for the simultaneous detection of YMV, YMMV and CMV. A total of 2616 plants in 65 locations were assessed in Ghana and 3583 plants in 75 locations in Nigeria. The incidence of virus symptoms across the locations was high with an average incidence of 94.1% for Ghana and 89.9% for Nigeria. The virus index detected YMV, YMMV and DBV. We also developed a RT-LAMP assay for YMV detection, a more sensitive and rapid nucleic acid detection technique. Further standardization is expected to result in simple and sensitive tool for YMV detection on-farm. Our work has also shown that 2% CTAB buffer is a relatively more efficient method for yam leaf sample preservation for diagnostic use. Sequence analysis suggested high homology among the RNA viruses (YMV and YMMV) and revealed that DBVs, integrated badnaviruses and endogenous pararetroviruses as the most prevalent in yams in Ghana and Nigeria. Results demonstrated high nucleotide diversity among badnaviruses, with possible evidence of the presence of new virus strains.

P0309: Microbes and Pathogens

Genome Assembly and Comparison of *Macrophomina phaseolina* Isolates on Strawberry and Alfalfa

Alyssa Burkhardt[1], Kevin Childs[2], Marina L. Ramon[1] and Frank Martin[1], (1)USDA-ARS, Salinas, CA, (2)Michigan State University, East Lansing, MI

Macrophomina phaseolina is a haploid fungus that typically causes charcoal rot on a wide range of hosts, but recently isolates with some level of strawberry host specificity have been identified. To understand this host specificity, we sequenced *M. phaseolina* isolates from strawberry and from alfalfa using PacBio and Illumina paired-end and mate pair platforms. The isolate from strawberry was assembled with FALCON and error corrected with PILON to yield 100 contigs with an N50 of 3.3 Mb. After Dovetail scaffolding, there are 84 scaffolds with an N50 of 4.3 Mb and a BUSCO completeness of 98% for the eukaryotic data set. The alfalfa isolate was also assembled with FALCON and has 27 contigs with an N50 of 4.2 Mb; Dovetail scaffolding resulted in 26 scaffolds and a similar N50. The BUSCO completeness for the fungal and eukaryotic datasets was 99%. Scaffold ends have been extended using MiSeq reads with SeqMan NextGen and some telomere sequences have been identified. A MAUVE and SynMap analysis of the two Dovetail scaffolded genomes show that the predicted scaffolds have several syntenic blocks, yet some chromosomal translocations and rearrangements were observed. Genome annotation is underway using the MAKER pipeline with RNA-Seq data which was collected under multiple conditions and assembled using Trinity. Additional comparative genomic analyses will be performed after the genome assemblies have been confirmed using BioNano optical mapping. The ultimate goal of this project is to compare the final genome assemblies and to identify regions that are associated with host specificity on strawberry.

P0310: Microbes and Pathogens

FANA_ASO Reduces Pathogens and Pest of Fruit Crops: Citrus and Grapevine

Wayne Hunter[1], **Jackie Lin Metz**[2,3], Michael J. Boyle[4], Greg McCollum[5], Veenu Aishwarya[3], Andres F. Sandoval Mojica[6] and Kristen Pelz-Stelinski[6], (1)USDA Agricultural Research Service, Fort Pierce, FL, (2)University of Florida, Ft. Pierce, FL, (3)AUM LifeTech, Inc., Philadelphia, PA, (4)Smithsonian Marine Station Fort Pierce, (5)USDA, Fort Pierce,, FL, (6)Univeristy of Florida, Citrus Research and Education Center, Lake Alfred, FL

FANA ASO's (Antisense oligonucleotides (ASO) modified with 2'-deoxy-2'-fluoro-D- arabinonucleic acid (FANA) were shown to reduce insect pests, their endosymbiotic bacteria, and a plant bacterial pathogen, Liberibacter asiaticus, Las. FANA provides a new non-transgenic strategy to manage agricultural pests and pathogens using sequence specificity and binding. Binding induces the cell natural Rnase H, cutting mechanism. Insects shown to be sensitive to FANA treatments included two hemipterans- Asian citrus psyllid, *Diaphorina citri*, glassy-winged sharpshooter, *Homalodisca vitripennis*, and the coleopteran *Diaprepes abbreviatus*, the citrus weevil. Adult insects fed on treated leaves or plants with FANA ASOs designed to each specific insect resulted in increased mortality. Psyllids 8 days post feeding 35%; Weevils at 15 days post feeding from 30% to 90%. Control scrambled, non-target FANA ASO, averaged 9% and water blank control resulted in 13 % mortalities. Endosymbionts in psyllid and leafhopper were significantly reduced after 8 days post feeding. Plant pathogen, Las, was reduced in infected citrus seedlings by 50% at 25 days post treatment. Delivery into tissues of insects and plants was visualized using confocal microscopy analysis fluorescence spectrophotometry. Fluorescently labeled FANA ASOs were observed in cells throughout the cell tissues: hemolymph, brain, fatbody, and alimentary tract, but not in eggs. The unique traits of FANA_ASO's provide a potentially new strategy to manage microbial

pathogens and their arthropod vectors for crop protection. Supported in-part: NIFA, USDA, Citrus Greening award #2015-70016-23028; and NIFA, USDA, award #2015-10479.

P0311: Microbes and Pathogens

Investigation of Liberibacter-Citrus Protein-Protein Interactions

Supratim Basu[1], Roel Rabara[1], Loan Huynh[2], Hau Nguyen[2], Jeanette Velasquez[2], Goutam Gupta[1], Shujian Zhang[1], Guixia Hao[3] and Ed Stover[3], (1)New Mexico Consortium, Los Alamos, NM, (2)NEW MEXICO CONSORTIUM, LOS ALAMOS, NM, (3)U. S. Horticultural Research Laboratory, USDA-ARS, Fort Pierce, FL

Huanglongbing (HLB) caused by *Candidatus* Liberibacter spp. is the most devastating disease of citrus. Like in other bacteria, Liberibacter effectors play an important role in the onset and progression of disease. However, in spite of the availability of the citrus and Liberibacter genome sequences, the molecular mechanism of interactions between Liberibacter effectors and citrus proteins still remains unclear. To better understand the roles of Liberibacter effectors, we have identified a putative effector protein (designated as $LasP_{235}$) of 123 amino acids with a eukaryotic nuclear localization signal (NLS) in the prophage region of the Liberibacter asiaticus (Las) psy62 genome. In an effort to identify the potential interacting partners of $P_{235,}$ we have overexpressed and purified the effector protein in *E. coli* (BL21). The purified, sequence verified protein was used for pull-down assays using total protein from healthy and Las-infected citrus leaves. The putative interacting complexes were purified using TALON resin and sent for identification by LC MS/MS, which helped identification of citrus proteins that were unique to the healthy, infected and some proteins that were expressed in both the healthy as well as the infected samples. Of these putative interactors we selected Aspartyl protease, glycosyl hydrolases, LTP (lipid transfer proteins) and SOD (superoxide dismutase) for further validation. The results from these experiments will be presented in detail. We believe that the results from these experiments will enable us to get an understanding of the pathogenicity mechanism of Las. Specifically, how the Las effectors manipulate the host defense responses. This understanding will lead to the development of HLB therapies.

P0312: Microbes and Pathogens

Genome Sequences of Two *Shewanella* Spp. Isolated from the Gut of the Sea Cucumber *Apostichopus japonicus* (Selenka, 1867)

Seongmin Cheon, Hyun-Hee Hong and Chungoo Park, Chonnam National University, Gwangju, South Korea

In this study, we sequenced the genomes of two Shewanella spp., newly isolated from the gut of the sea cucumber Apostichopus japonicus (Selenka, 1867). The whole-genome sequences reported here will expand the repertoire of genomic information for the members of the genus Shewanella and will provide important insights into their roles within microbial communities.

submitter's E-male Adress: 141855@jnu.ac.kr

P0313: Microbes and Pathogens

Draft Genome Sequences of *Pseudoalteromonas tetraodonis* CSB01KR and *Pseudoalteromonas lipolytica* CSB02KR, Isolated from the Gut of the Sea Cucumber *Apostichopus japonicus*.

Jooseong Oh, Jihoon Jo, Sung-Gwon Lee and Chungoo Park, Chonnam National University, Gwangju, South Korea

We present here the complete genome sequences of two newly isolated *Pseudoalteromonas tetraodonis* and *Pseudoalteromonas lipolytica* strains, isolated from the gut of the sea cucumber *Apostichopus japonicus*, to provide a useful means for facilitating the study of antibacterial, bacteriolytic, agarolytic, and algicidal activities of marine *Pseudoalteromonas* species.

P0314: Microbes and Pathogens

Investigation of the Massive Gene Expression for the Active Growth Conditions for the Symbiotic Bacteria in Entomopathogenic Nematodes

Yeon Su Yu, Daedong Tech Co, Gyeongsan, South Korea, Krishna Bahadur Karki, Yeungnam University, Gyeongsan, South Korea and **Kwang-Hyun Baek**, Department of Biotechnology, Yeungnam University, Gyeongsan, South Korea

Entomopathogenic nematodes in the genera of *Steinernema* and *Heterorhadditas* can be used for effective biological control agents against insects, but they are non-toxic to vertebrates including human being. Currently the entomopathogenic nematodes is extensively used in many developed countries and the second most highly traded group of biocontrol products. Recent great leap for the mass production has been possible using solid-culture system instead of traditional liquid culture system. One of the most important factors for the success of the solid-culture system is inoculation of the symbiotic bacteria *Xenorhabdus nematophils* and *Photorhabdus luminescens* in the nematodes, however, there have been problems of successive incubation and long-term storage of the bacteria. Recent finding of a correct medium significantly increased the growth and long-term storage of the bacteria, therefore, the massive gene expression revealed the key regulatory genes expressed by the culture media. In this presentation, the current status of the technology, research in Korea and future application of entomopathogenic nematodes along with the genes responsible for the active growth of the symbiotic bacteria are discussed.

P0315: Microbes and Pathogens

Major Transcriptional Reprogramming of *Phytophthora parasitica* Prior to Infection in Citrus

Ronaldo JD Dalio, Centro de Citricultura/IAC, Cordeirópolis - SP, Brazil, Heros Maximo, Biotechnology Lab, IAC, Cordeirópolis, Brazil and Marcos A. Machado, Sylvio Moreira Citrus Center, Agronomy Institute, Cordeiropolis, Brazil

Brazil and USA are the top orange producers in the world, however, there is a constant threat from diseases, such as root rot caused by *Phytophthora parasitica*. It is well known that *Phytophthora* species secrete hundreds of effectors that enable infection. Manipulating effectors can lead to pathogen loss of virulence. The mechanistic molecular functioning of effectors remains poorly understood for phytophthora-citrus diseases. We were interested in the moments before infection, i.e. how *P. parasitica* prepares itself facing signals that plants are around, also how many genes are differentially expressed and which one are putative effectors? Are there differences in transcription reprogramming of *P. parasitica* facing signals from a susceptible or a resistant host? To answer these questions we cultivated *P. parasitica* in liquid culture with or

without root extracts of a susceptible citrus rootstock (*Citrus sunki*) and a resistant one (*Poncirus trifoliata*). We harvested the mycelia to perform a RNAseq through Illumina sequencing. Surprisingly, there are much more down-regulated genes than up-regulated ones in both treatments with the susceptible or resistant plant extracts in comparison with control treatment. However, among the up-regulated genes, we have found several apoplastic and citoplasmatic effectors, such as RxLR and Crinkler effectors (also plotted in the genome architecture of *P. parasitica*). There were much more up-regulated effectors in the mycelia treated with the sunki (susceptible variety) extract in comparison with the ones treated with *P. trifoliata* extracts, suggesting that the pathogen can recognize different hosts and activate a distinct arsenal of effectors prior to infection.

P0316: Microbes and Pathogens

Towards Understanding the Parasitic Nature of the Reniform Nematode using Microbiome and Metagenome Analyses

Grace Taylor[1], Zachary B. Gossett[1], Abreeotta Williams[2], Seloame T. Nyaku[3], Kathy Lawrence[4], Govind C. Sharma[1] and Venkateswara R. Sripathi[1], (1)Alabama A&M University, Normal, AL, (2)Columbia State Community College, Columbia, TN, (3)University of Ghana, Accra, Ghana, (4)Auburn University, Auburn, AL

Plant-parasitic nematodes are a threat to global food security with an estimated annual loss of over $100 billion. Among these, *Rotylenchulus reniformis* is one of the major pests in the Southern United States with a broad host range of over 200 crop plants including Upland Cotton (*Gossypium spp.*). Our genome and transcriptome analysis of the reniform nematode (RN) revealed several parasitism-related genes that are of fungal and bacterial origin. However, the genetic basis for pathogenicity of the female RN is not elucidated yet. This study aimed at understanding parasitic nature of the pest by i) comparing the microbiomes of both male and female RN, and ii) screening the bacterial diversity in metagenome samples collected from heavily-infested, moderately-infested and low-infested cotton fields in Alabama to understand effects of the soil type on nematode-associated fungi and bacteria. The samples were collected from selected sites as per the Alabama extension protocol for soil collection for nematode analysis from row crops, and included a minimum of 30 individual samples. Approximately, 25 single females and males were selected for microbial DNA and RNA extraction (ZymoBIOMICS). Microbial diversity assessed by sequencing single-end libraries of amplified 16S rDNA genes from microbiomes of male and female RN. The reads are clustered into operational taxonomic units (OTUs) and phylogenetic trees were constructed with closely related bacterial species to compare microbiomes. Our preliminary results indicated that majority of OTUs identified belonged to the phylum Proteobacteria. Separately metagenome analysis was conducted from three selected field sites representing varying levels of RN infestation and sequenced in triplicates. Furthermore, this study also aimed at screening the microbiomes of foregut, midgut and hindgut regions of the RN using microdissection, which will aid in understanding the microbial specificity to these sections and the genetic basis for the parasitism of RN.

P0317: Microbes and Pathogens

Intraspecific Diversity of Recombination Rate and Interference in *S. cerevisiae*

Xavier Raffoux[1], Mickael Bourge[2], Olivier C. Martin[1] and **Matthieu Falque**[1], (1)GQE– Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Gif-sur-Yvette, France, (2)Cytometry / Electronic Microscopy / Light Microcopy Facility, Imagerie-Gif, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France

Meiotic recombination is a major driver of genome evolution, and more insights into the genetic control of crossing-over would be beneficial for many fields of fundamental and applied genetics. We used a high-throughput method to measure recombination rate and crossover interference in 26 *S. cerevisiae* strains representing a large part of the diversity of the species. Sixteen intervals were monitored, covering chromosomes VI and XI entirely, and part of chromosome I. Average recombination rates ranged between 0.18 and 0.52 cM/kbp across strains, and specific intervals showed up to 9.5-fold variation. Recombination landscapes along chromosomes also varied between strains. In some adjacent intervals, we observed positive interference which varied across strains and was positively correlated with crossover number. Recombination rate was negatively correlated with sequence divergence between homologs at the whole genome scale, in whole intervals, and in DSB rich regions within intervals. The results suggest that these correlations may not be explained by cis-effects only. Finally, to obtain more insights on the genetic architecture of crossover rate, we built an incomplete diallel design and measured recombination rates in one region of chromosome XI for 10 different hybrids obtained by crossing five parental strains. The results suggest that recombination rate across hybrids in this design was mainly controlled by the level of sequence divergence between parental strains and also by inbreeding effects in homozygotes. In contrast, other interaction effects, and additive effects were weak.

P0318: Microbes and Pathogens

The Exceptionally Large Genome of the Harmful Red Tide Dinoflagellate *Cochlodinium polykrikoides* Margalef (Dinophyceae): Determination by Flow Cytometry

Hyun-Hee Hong, Jihoon Jo and Chungoo Park, Chonnam National University, Gwangju, South Korea

Cochlodinium polykrikoides is a red-tide forming dinoflagellate that causes significant worldwide impacts on aquaculture industries and the marine ecosystem. There have been extensive studies on managing and preventing *C. polykrikoides* blooms, but it has been difficult to identify an effective method to control the bloom development. There is also limited genome information on the molecular mechanisms involved in its various ecophysiology and metabolism processes. Thus, comprehensive genome information is required to better understand harmful algal blooms caused by *C. polykrikoides*. We estimated the *C. polykrikoides* genome size using flow cytometry, with detection of the fluorescence of DNA stained with propidium iodide (PI). The nuclear genome sizes of *C. polykrikoides* were 100.97 and 110.54 Gb, as calculated by comparing their mean fluorescence intensity (MFI) to the MFI of *Mus musculus*, which is 2.8 Gb. The exceptionally large genome size of *C. polykrikoides* might indicate its complex physiological and metabolic characteristics. Our optimized protocol for estimating the nuclear genome size of a dinoflagellate using flow cytometry with PI can be applied in studies of other marine organisms.

P0319: Microbes and Pathogens

Genome Analysis of High Light Tolerant Chlorella Strains

Lena Altrogge[1], Christian Pfaff[2], Janka Widzgowski[2], Alexander Vogel[3], Bjoern Usadel[4], Ulrich Schurr[4] and Heiko Schoof[1], (1)INRES Crop Bioinformatics, University of Bonn, Bonn, Germany, (2)Forschungszentrum Jülich, IBG Plant Sci 2, Jülich, Germany, (3)Institute for Biology I, RWTH Aachen University, Aachen, Germany, (4)Forschungszentrum Jülich GmbH, Jülich, Germany

Different green algae *Chlorella* strains from Arctic regions, temperate latitudes, and subtropical climate were reported to grow optimally even at light intensities above 1,500 µmol photons m-2s-1. We performed genome annotation of four *Chlorella* strains with the genome annotation pipeline BRAKER1 and observed differences between the high light sensitive strain *C. vulgaris* 211-11b and the high light tolerant strains in gene counts and gene density. Groups of orthologous genes revealed close phylogenetic relationship among the analysed *Chlorella* strains whereas fewer conserved orthologs and more strain specific genes were identified in the reference genome of *Chlamydomonas* reinhardtii. While 40% of all orthologs were conserved in all *Chlorella* strains, we observed a large proportion (30%) of orthologous groups specific to high light tolerant strains revealing considerable differences in the genomes of high light tolerant strains compared to the high light sensitive strain. We analysed high light specific proteins and identified processes possibly affecting high light resilience.

P0320: Microbes and Pathogens

What's the Difference between Pathogenic and Non-Pathogenic Oomycetes? – a Transcriptomic Profiling of the Marsh Grass-Decomposing Oomycete Salisapilia

Sophie de Vries, Jan de Vries, John M. Archibald and Claudio H. Slamovits, Dalhousie University, Halifax, NS, Canada

Oomycetes encompass many of the most devastating pathogens of plants and animals. Genomes and transcriptomes of these pathogens have been well established over several years of research and mined for virulence factors. But what are the differences between pathogenic and non-pathogenic oomycetes? Here we study *Salisapilia sapeloensis,* an oomycete isolated from and suspected to decompose marsh grass litter. We developed an axenic system to grow *Salisapilia* with and without marsh grass litter and confirmed its interaction with the litter macro- and microscopically. RNA-Seq was performed from *Salisapilia* grown in rich medium (control), marsh grass litter and the water surrounding the litter. Initial analyses showed a strong opposite reaction in control vs. marsh grass-decomposing *Salisapilia*: of the 8023 genes 24.6% (1971 genes) were differentially expressed in the two treatments. Comparing marsh grass-decomposing *Salisapilia* and *Salisapilia* in close association to the litter, 3.1% (250 genes) were significantly expressed. The overlap between the two datasets corresponds to 245 differentially regulated genes, signifying genes specific to marsh grass-decomposition. A third of these were unknown, as is often the case for oomycete pathogens due to their many effector genes, but remain to be characterized for a non-pathogenic oomycete. Considering only the 10 most up-regulated genes from hits corresponding to oomycetes, most of them encode proteins with domains important for lytic enzymes, for example pectate lyases, cellulases and lipid esterases. Our data will help to pinpoint the distinction between an oomycete pathogen infecting plant tissue and an oomycete that simply decomposes it.

P0321: Microbes and Pathogens

Intra-Specific Comparison of Mitochondrial Genome in *Tremella fuciformis* Suggests a Model of Host Gene Fragment Exchange *via* Intron Mobility

Youjin Deng, Fujian agriculture and forestry university, Fuzhou, China, Baogui Xie, Fujian agriculture and forestry university, fuzhou, China and Ray Ming, University of Illinois at Urbana-Champaign, Urbana, IL

Mitochondrial genome comparisons among closely related population may help to investigate their evolution mechanism. In this study, whole genome sequencing of 16 *Tremella fuciformis* isolates (TF01-TF16) was performed using Illumina and/or Pacbio sequencing technologies. Mitochondrial DNA (mtDNA) related contigs were extracted and assembled into finished circle molecules, ranging from 35,104 bp to 49,044 bp in size. Intra-specific comparison of *T. fuciformis* revealed that all mtDNAs contained the same set of 41 conserved genes, and shared the same gene order except for that of TF08. A 9517-bp fragment containing six conserved genes in TF08 was reversed through homologous recombination. Comparative analyses also showed that introns and intergenic regions were the two most variable regions. Among 24 introns detected, eleven located to protein-coding genes, three in tRNAs, and other ten in rRNAs. In addition, two mobile fragments were found in intergenic regions. Interestingly, six structures of gene with N-terminal duplication were found to result from insertion of intron with truncated host gene precursor. Comparison of genes with and without N-terminal duplication gave rise to a possible model of N-terminal gene fragment exchange through gain or loss of an intron with a truncated host gene precursor. The model suggested a new gene evolution approach of mitochondrial genes: partial gene fragment was replaced by exogenetic one via intron mobility.

P0322: Natural Populations

Transcriptomics of Adaptation in Non-Model Species

Gregory Maes, Centre for Human Genetics | Genomics Core Leuven, Leuven, Belgium

Cichlid fishes from the Great African lakes are often considered as the ultimate models to investigate adaptive radiation and speciation. We investigated the genetic basis of the cichlid speciation process itself through a dual analytical approach: comparing the genome wide differentiation within and between sister species exploring the transcriptomic basis of the behavioural aspects (display and acoustic signalling during courtship) allowing female fishes to differentiate between conspecific and hetero-specific males.

Our objectives were (1) to identify the genes involved in the courtship process by detecting up- or downregulated genes during mate pairing compared to the control condition and (2) to investigate whether a differential gene expression is observed when females are confronted with a hetero- or a conspecific male. Results show a clear transcriptional variation between brain regions in cichlids, and a significant difference in gene expression between con- and heterospecific interaction before and after egg laying.

P0323: Natural Populations

Genetic, Geographic, and Climate Diversity of the Brachypodium distachyon Species Complex

Jared C. Streich[1], Pip B. Wilson[2], Steven Eichten[3], Shuangshuang Liu[4], Kevin D. Murray[2], Diana López-Alvarez[5], Hikmet Budak[6], Pilar Catalan[7], Luis Mur[8], Kent Bradford[9], David F. Garvin[10], Smadar Ezrati[11] and Justin O. Borevitz[2], (1)Australian

National University, Canberra, ACT, Australia, (2)The Australian National University, Canberra, Australia, (3)Australian National University, Canberra, Australia, (4)University of California, Davis, Davis, CA, (5)University of Zaragoza, Huesca, Spain, (6)Montana State University, Bozeman, MT, (7)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain, (8)Aberystwyth University, Aberystwyth,, United Kingdom, (9)University of California, Davis, CA, (10)USDA-ARS, PSRU, St. Paul, MN, (11)Tel-Aviv University, Tel Aviv, Israel
The now globally distributed *Brachypodium distachyon* species complex (two diploids, *B. distachyon* and *B. stacei*; and their allotetraploid *B. hybridum)* are an ideal invasion biology system to study: migration, habitat vulnerability, climate breadth, and genetic diversity across geography. We genotyped by sequencing 1,897 individuals from 458 global locations, identified species via genome-specific markers, calculated population structure, allelic richness across geography, climate diversity, and abundance. We found 129 genotypes using 14,436 SNPs from 479 *B. distachyon,* 13 genotypes using 4,744 SNPs from 50 *B. stacei,* and 80 genotypes from 1,015 *B. hybridum* using 18,525 variants. Besides one Australian location, *B. distachyon* was Mediterranean exclusive like *B. stacei*. *B. hybridum* was found globally. We computer modelled global suitability per species to search possible new native sampling locations, and vulnerable habitats globally. Across South America, South Africa, Australia, and North America 2,770,680 km^2 were classified *B. hybridum* vulnerable. We permutation tested *B. hybridum* genotypes in abundance and climate diversity having nine or more independent occurrences and presence in 3-plus climate classes. Many *B. hybridum* genotypes were found in non-native ranges, having wide climate windows or global occurrence, but genotype NRD-1 had significantly larger of both. NRD-1 locations were used to computer model potential range, finding that 18,114,261 km^2 were sensitive globally at the lowest calculated suitability score 0.05. Thus, we conclude genotype NRD-1 is more prone to dispersal and has a significantly larger climate breadth than random.

P0324: Natural Populations

Population Structure of Wild and Cultivated Plants Shows Hierarchical Organization of Cassava Germplasm Diversity.

Fausto Villafrade Rodriguez Zapata[1], Tatiana Ovalle[1], Manuel Ruiz[2], Anestis Gkanogiannis[3] and L. Augusto Becerra Lopez-Lavalle[4], (1)CIAT, Cali, Colombia, (2)CIRAD, UMR AGAP, Montpellier Cedex 5, France, (3)CIRAD, UMR AGAP, Montpellier, France, (4)International Center for Tropical Agriculture, CIAT, Cali, Valle del Cauca, Colombia
Cultivated cassava diversity has been shaped and is influenced by its wild relatives gene pools. Its primary gene pool (GP-1) consists of the flabellifolia-peruviana subspecies complex that cross easily with domesticated plants. Its secondary gene pool (GP-2) spans Manihot wild species that have a limited ability to interbreed with cassava. RAD sequencing of 640 accessions from the core collection at CIAT, with increased sampling of GP-1 and GP-2, shows a kinship connectivity of 9% and 175 nonrelated individuals. GP-2 sources of cultivated material include but are not limited M. glaziovii. A clear subdivision of the flabellifolia complex is not congruent with the previously postulated peruviana-flabellifolia-tristis taxonomic division. At least 5 different genetically isolated populations around the Amazon Basin could be founding stocks of current cultivated Manihot esculenta. Phylogenetic and pedigree analysis show a hierarchical structure of the populations in Latin America and the Caribbean in both wild and cultivated cassava with extensive gene flow and admixture. This hierarchical structure is dominated by kinship relations, with 95% of the sequenced samples having a third degree relative, and 75 % belonging to a single 1st degree cluster. We propose that this pedigree structure most likely is the result of modern breeding programs carried out by institutions like CIAT and EMBRAPA. Nevertheless strong phylogenetic signal can be detected from the unrelated accession set that could go back to the columbian exchange and holocene domestication. Consequences on breeding and the study of the domestication process are further discussed.

P0325: Natural Populations

Natural Variation of Leaf Secondary Metabolites and Underlying Genetics in European White Oaks

Benjamin Brachi[1], Grégoire Leprovost[1], Stéphane Bernillon[2], Céline Lalanne[1], Annick Moing[2] and **Christophe Plomion**[1], (1)INRA, Université de Bordeaux, Cestas, France, (2)INRA, Université de Bordeaux, Bordeaux, France
Trees are major components of the forest ecosystem and are the center of complex interactions with herbivores, microbes and other plants. A major mechanism by which plants interact with this environment is the production of secondary metabolites. Oaks produce a complex blend of secondary metabolites and the ecological effects of these compounds have long been studied. However, the natural variation of secondary metabolites in forest trees has received less attention. In this study we explore variation of leaf secondary metabolite production in European white oaks, using high throughput mass spectrometry in three tree collections: (i) a family of 204 *Quercus robur* full-sibs allowing quantitative trait loci analysis, (ii) individuals from nine European provenances of *Q. petraea* grown in a common garden, and (iii) eight natural populations of *Q. petraea* located along an elevation gradient in the French Pyrenees. We screened close to1800 samples and detected approximately 130 leaf secondary compounds. We found that leaf secondary compounds variation is to a large extent determined by few loci of large effects. We also found surprisingly little between population phenotypic differentiation, and high levels of within population variation. Interestingly, certain compounds segregate at intermediate frequencies in all provenances investigated suggesting these polymorphisms are maintained at balanced frequencies. In ongoing analysis we explore candidate genes and signature of selection at the loci underlying this variation.

P0326: Natural Populations

Analysis of Transposable Element Activity and Density within Natural Populations of Arabidopsis.

Elizabeth, R. Waters, San Diego State University, San Diego, CA
Transposable elements (TEs) are found in all genomes. However, the proportion of each genome that is comprised of TEs can vary significantly across species. Transposable element density and activity can also vary within species. Here we report on our study of the density and activity of TEs in 16 populations of *A. thaliana* that are found along an altitudinal gradient. Populations surveyed range from sea level to 1,700 meters in elevation. Previous studies of these populations have demonstrated significant differences in tolerance to heat stress between the high and low elevation populations as well as fitness and life history trait variation among these populations. It has been reported that some transposable elements are activated by heat stress and we sought to determine if transposable element activity and density also varies across these populations and if this variation is correlated to the frequency of heat stress experienced by these populations. Complete genome sequence of 32 individuals from 16 populations were generated and analyzed. We identified novel insertion sites compared to the reference *A.*

thaliana Col-0 genome for all individuals. We found significantly more TE activity, as defined as new insertion sites, in the high elevation populations. Notably, the activity of TEs among populations also varied by TE family. The *Copia* and *Helitron* families were the most active across all populations. We can conclude that TE activity is correlated with climatic variables and suggests that the high elevation populations are experiencing more genomic stress.

P0327: Natural Populations

Analyses of Fungal Microbiomes Associated with Plants Endogenous to Central Texas

Charles Hauser and Jacquelyn Turcinovic, St. Edward's University, Austin, TX

Microbial communities in plant roots provide critical links between above and below ground processes in terrestrial ecosystems. The composition of root-associated microbial communities has been attributed to various factors including: plant and microbe genotypes, plant exudates, soil microbial composition, and soil characteristics. Bacterial assemblages are known to facilitate plant uptake of critical nutrients such as phosphorus and nitrogen, promote plant growth, and both mediate protection from and sensitivity to pathogens. However, the composition and contributions of fungal microbiomes to plant species in native environments are less well characterized. The fungal communities associated with roots of eight species plant species native to central Texas were characterized including: *Schizachyrium scoparium, Arbutus xalapensis, Muhlenbergia reverchonii, Nolina lindheirmeriana, Prosopis glandulosa, Yucca rupicola, Juniperus ashei,* and *Carex planostachys*). DNA was extracted and 18S rDNA (ITS1-ITS2) was sequenced from four soil fractions associated with each plant species: soil outside the rhizosphere not penetrated by plant roots (bulk); soil surrounding the rhizosphere and below the O-Horizon (neighboring); one to two millimeters of soil touching the root and root hairs (rhizosphere); and within plant tissues (endosphere). Principle component analysis reveals a clustering of endosphere samples vs non-endosphere samples, thus potentially identifying "within-plant" from "outside-plant" fungal populations. The endosphere fractions of five of the eight plants appear enriched for class Sordariomycetes (Ascomycota) which was not restricted to monocots or eudicots. Additional analyses are currently in progress, including a more detailed examination of alpha diversity and network structure.

P0328: Natural Populations

Extrinsic Forcing of Genomic Evolution during Speciation: A Geo-Genomic Study of Gopherus Desert Tortoises

Greer A. Dolby, Timothy H. Webster, Dale F. DeNardo, Melissa A. Wilson Sayres and Kenro Kusumi, Arizona State University, Tempe, AZ

There are often multiple extrinsic forces shaping patterns of genomic evolution among populations or species, and these forces operate on varying timescales. To understand how threatened *Gopherus* tortoises have evolved in response to the changing environments of the southwestern US requires accounting for the suite of extrinsic (geo-climatic) forces to which they have responded. Yet rarely are primary geologic and climatic data rigorously integrated into genomic studies in order to disentangle these relative drivers of organismal evolution. Because genomic divergence and speciation can occur by neutral drift, differential adaptation, or both, utilizing statistical frameworks in which multiple, non-mutually-exclusive hypotheses can be tested is necessary. Here we present ongoing work using low-coverage (~5x) whole-genome sequencing of individuals from sister *Gopherus* species and their hybrid populations. We aim to better understand the historic speciation and ongoing/recent hybridization of these non-model species through reconstructing demographic histories of both lineages, determining regions of elevated interspecific divergence, testing for genic regions under selection, and assessing signatures of genomic introgression. In addition, we are using primary sediment, tectonic, and climatic data over the past ~8 million years to establish geo-climatic hypotheses regarding what extrinsic factors might have shaped these patterns of genome evolution. Specifically, geologic development of the lower Colorado River region and paleo-monsoon system would likely have mediated gene flow and differential adaptation, respectively. Using a geo-genomic approach to disentangle the relative contributions of these extrinsic processes to genomic divergence in these species would advance our understanding of speciation and drivers of genomic evolution broadly.

P0329: Natural Populations

The Genomic Complexity and Diversity of a Chromosome-Wide Inversion in a Songbird

Vinicius H. da Silva[1,2], Veronika N. Laine[2], Mirte Bosse[1], Kees van Oers[2], Martijn F.L. Derks[1], Marcel E. Visser[2], Richard P.M.A Crooijmans[1] and Martien A.M. Groenen[1], (1)Wageningen University & Research Center, Wageningen, Netherlands, (2)The Netherlands Institute of Ecology (NIOO), Wageningen, Netherlands

Chromosomal inversions represent an important class of polymorphisms that are of particular interest in evolutionary studies. Inversions are found in various bird species and are associated with traits related to sexual behavior and sperm motility. Great tit (*Parus major*) is a songbird that is extensively used as a model in ecology and genetics. Using PCA clustering followed by FST on ≈500k SNPs from 2,296 birds, we identified a pericentric inversion overlapping ≈90 % of chromosome 1A (64.2 Mb). This inversion is present in a heterozygous state in 5 % of the population (117 birds). We identified a 60 kb copy number gain that is associated with the inversion and located close to the tentative downstream breakpoint of the inversion. This copy number gain (≈10 copies), validated using whole genome sequencing data, results in an increase of the size of chromosome 1A of ≈600 kb. The allele distribution in the inverted phase differs throughout the chromosome, suggesting some degree of recombination/mutation in a block with approximately 30 Mb in the middle of the inversion. As recombination is unexpected between inverted and normal phases, more recent haplotypes of the inversion may have a genomic interval in the middle which is in the normal orientation. However, further analyses are required to clarify existing haplotypes, structural complexity and the biological effect of the inversion.

P0330: Natural Populations

First Draft Genome Sequence of Anadromous Hilsa Shad (*Tenualosa ilisha)* and Development of Genomic Resources for Conservation

Md Bazlur R. Mollah, Md Golam K. Khan, Md Shahidul Islam and Md Samsul Alam, Bangladesh Agricultural University, Mymensingh, Bangladesh

Hilsa shad *(Tenualosa ilisha),* belongs to the order *Clupeiformes*, an anadromous species lives in the lower region of the estuaries and the foreshore areas, ascends the rivers during the breeding season and return to the original habitat after spawning. The fish is found in the Bay of Bengal, Persian Gulf, Red Sea, Arabian Sea, Viet Nam Sea and China Sea. The riverine habitat covers the Padma, Jamuna, Meghna, Karnafully and other coastal rivers of Bangladesh, the Satil Arab, and the Tigris and Euphrates of Iran and Iraq, the Indus of Pakistan, the rivers of Eastern and Western India and the Irrawaddy of Myanmar. Here, we describe the first draft whole genome sequence of an anadromous species belonging to the *Tenualosa* genus, based on material collected from the Bay of Bengal and Meghna river. The draft genome consists of 712,030,140 reads, which yielded an assembly of 2,321,352 contigs and 249,092 scaffolds. The estimated genome size was 768 Mb and an average G+C content was 43.8%. We also identified 9,867 SSRs and 25,782 SNPs. The release of the genome sequence of hilsa and the developed genomic resources might contribute to the understanding of the underlying physiology, reproductive behavior and conservation of an anadromous species.

P0331: Natural Populations

Engineering Synthetic Speciation to Control Invasive Carp: Simulation with Agent Based Model

Siba R. Das[1], Maciej B. Maselko[1], James E. Parker[2], Przemyslaw G. Bajer[3] and Michael J. Smanski[1], (1)Department of Biochemistry, Molecular Biology, and Biophysics and BioTechnology Institute, University of Minnesota, St. Paul, MN, (2)Computer Science and Engineering, University of Minnesota, Minneapolis, MN, (3)Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN

Aquatic invasive species such as the carp are widely distributed in the United States, mainly to the mid-western region including the Great Lakes. In addition to destroying the habitat for native species in the region by competing with and depleting the resources, they impart economic loss by interfering with the commercial fishing activities. An estimated $7 billion every year of commercial activities are at risk from habitat destruction created by carp in the Great Lakes basin alone. Control measures so far have been focused on physical methods such as, netting, poisoning and creating physical barriers, which have been mainly ineffective so far. Using genetic engineering we have developed a rather radical approach to create a species like barrier. This barrier can effectively control population of the invasive carp leading to extinction over time. Using an Agent based model we show here that the control of invasive carp with our method is substantially better than existing methods. We also show that our method outperforms other genetically control measures described elsewhere, such as female lethality and gene drive systems because of its robustness to resist unintended genetic changes in the system. We acknowledge funding by DARPA and University of Minnesota internal grant.

P0332: Natural Populations

Understanding Predator – Prey Dynamics from Non-Invasive Camera Trapping, Diet and Individual Identification using Fecal DNA

Jordana Meyer, Kevin Leempoel and Elizabeth Hadly, Stanford University, Palo Alto, CA

Rapid global urbanization poses a tremendous threat on the composition of animal communities, as they are forced to evolve, or dissolve, amongst these fragmented habitats. These changes can impact complex intra – and interspecies relationships, causing further cascading effect over time. At the Jasper Ridge Biological Preserve, CA, a reserve surrounded by residential areas, long-term camera trapping efforts (2009 – 2017) show a remarkable predator – prey cycle with four predator species, puma *(Puma concolor)*, bobcat (*Lynx rufus)*, coyote (*Canis latrans)*, gray fox (*Urocyon cinereoargenteus*) and two main prey species, black-tailed deer (*Odocoileus hemionus*) and Black-tailed jackrabbit (*Lepus californicus*). However, the role of predation, weather, and anthropogenic pressure cannot be easily disentangled from the camera trap analysis. In this project, we propose to collect scats in order to determine the predator – prey relationship and the influence of an apex predator on niche breadth and partitioning in a carnivore guild using DNA metabarcoding and Next Generation Sequencing. We will first determine predator and prey diet using barcoding genes. Second, we will estimate the population size and spatial distribution of these four key predators through a genetic capture-recapture approach, based on SNP arrays and microsatellites from the same fecal samples. And finally, we will look at individual diet preference within the four predator species. To better understand the effects of human pressures on predator-prey dynamics, we propose to conduct a comparative study within the urbanized buffer zone of Jasper Ridge. Most importantly, this approach could then be applied in wilderness spaces globally.

P0333: Equine

EquCab3: A New Horse Reference Genome Assembled Using First-, Second-, and Third-Generation Sequencing Data

Edward Stallknecht Rice[1], Michael S. DePriest[2], Ludovic Orlando[3], James N. MacLeod[4] and Theodore S. Kalbfleisch[2], (1)UC Santa Cruz, Santa Cruz, CA, (2)University of Louisville, Louisville, KY, (3)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, (4)Department of Veterinary Science, University of Kentucky, Lexington, KY

The current reference assembly for the domestic horse *Equus caballus* was published in 2009. This assembly used Sanger (first-generation) sequencing along with bacterial artificial chromosomes (BACs) to produce an assembly with 112kb contig N50 and 46Mb scaffold N50. Since 2009, there have been enormous advances in sequencing technology, bringing the cost and time required for sequencing down by more than a factor of 10,000. We used the existing Sanger sequence data along with several of these new technologies to assemble a new reference genome. Illumina HiSeq short reads increased average assembly read depth six-fold, resulting in fewer mis-calls. We used CHiCago and Hi-C long-insert libraries to improve scaffold assembly, nearly doubling the scaffold N50 and increasing the amount of sequence assigned to chromosomes. Gap-filling with PacBio long reads greatly increased contig sizes. We used a 10x Chromium library to identify and phase variants. The final assembly has 4,493kb contig N50, 85Mb scaffold N50, and 70Mb more sequence assigned to chromosomes. It will be publicly available by the end of 2017. This poster will discuss the various approaches and software pipelines we used to leverage the strengths of our different data types to create the best possible final assembly, including those that did not work, with the hope that this information will be useful to those producing their own reference genomes using these new technologies.

P0334: Equine

Strategies for Finding Previously Unannotated Protein-Coding Genes in an Updated Mammalian Reference Genome, Featuring EquCab3

Michael S. DePriest[1], James N. MacLeod[2] and Theodore S. Kalbfleisch[1], (1)University of Louisville, Louisville, KY, (2)Department of Veterinary Science, University of Kentucky, Lexington, KY

As sequencing technology continues to progress, reference genomes assembled using legacy technology are constantly re-evaluated using new types of sequence data. New sequencing methods allow researchers to fill gaps and fix errors, resulting in better, more complete reference genomes. In the case of EquCab3, the forthcoming new reference genome for the domestic horse, *Equus caballus,* the filled gaps contain many genes of immediate interest to the equine research community. The presence of protein-coding genes in newly characterized regions present in EquCab3 has been demonstrated previously using a BLAST-based approach. However, due to similarities among protein sequences in the database, the resolution of this analysis was limited and the results contained many false positives. In the current study, we evaluated different strategies for finding genes in a mammalian genome assembly using EquCab3 as the primary example. We investigated the effect of different experimental designs, including changing score thresholds, masking the genome prior to searching, and searching transcripts rather than the genome sequence. We used RNA-seq data to validate our gene predictions and determine gene structure. Finally, we compared our results from the horse genome with other recently updated mammalian reference genomes.

P0335: Equine

Tracking Six Millenia of Horse Selection, Admixture and Management with Complete Genome Time-Series

Antoine Fages[1,2], The ERC PEGASUS consortium[2] and Ludovic Orlando[1,2], (1)Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, (2)Laboratoire d'Anthropologie Moléculaire et d'Imagerie de Synthèse, CNRS UMR 5288, Université de Toulouse III Paul Sabatier, Toulouse, France

The domestication of the Horse and its impact on warfare, transportation and agriculture, have revolutionized human history. Even though most modern breeds have been engendered within the last couple of centuries, humans have managed horse livestock for over five millenia. Recent selective and management strategies have tremendously impacted the genetic structure of horse populations. As a result, modern patterns of genetic diversity can only partly help reconstruct the horse domestication process prior to the modern era. Recent research in our laboratory, carried out in the framework of the ERC PEGASUS programme, has endeavoured to sequence complete horse genomes from accross their whole temporal and geographical domestication range in order to identify how the many past human cultures progressively forged the horse genome by means of selection, drift and admixture. This work revealed two different dynamics at play within early and late domestication stages, involving the selection for different functional pathways, different management strategies for the genetic resource available, including stallion diversity, and a recent increase in the genomic deleterious load. Our new genome dataset now allows us to document such changes at unprecedented scales and reveals unexpected features of the whole population dynamic underlying horse domestication.

P0336: Equine

De Novo Assembly of the Equine MHC Region Using Linked-Read Genome Sequencing

Donald Miller[1], Peter Schweitzer[1], Theodore S. Kalbfleisch[2], James N. MacLeod[3], Ernest Bailey[4] and Douglas Antczak[1], (1)Cornell University, Ithaca, NY, (2)University of Louisville, Louisville, KY, (3)Department of Veterinary Science, University of Kentucky, Lexington, KY, (4)University of Kentucky, Lexington, KY

The Major Histocompatibility Complex (MHC) has been associated with more disease conditions than any other region of the equine genome, including susceptibility to equine sarcoid tumors, uveitis, insect bite hypersensitivity, and abortion. However, sequencing and assembly of the MHC presents a tremendous challenge in genomics because of the region's high degree of polymorphism, gene duplication, and structural variation between MHC haplotypes. Because of these features, it is not possible to assemble conventional short read (e.g. Illumina) sequencing data into reliable MHC region haplotypes. Here we produced a *de novo* assembly of the equine MHC using the 10x Genomics™ Chromium™ linked read gel-bead system. For input DNA we used Twilight, the DNA donor of the NCBI horse genome sequence, and compared our results to the EquCab2 assembly. We obtained 120 GB of sequence (~50x coverage) that was assembled *de novo* at the Cornell Biotechnology Resource Center using the 10x Genomics Supernova™ assembler program. This produced six long contigs in the MHC region with very few gaps, allowing us to correctly order the class I and class II genes on the ELA-A3 haplotype and to obtain high fidelity full length genomic sequences for all of those genes. The 10X Genomics™ platform produced data that is virtually identical to that found in the current NCBI EquCab2 equine genome assembly of the MHC region. These technologies should be useful for assembly of other equine MHC haplotypes, and for investigating the mechanisms of MHC disease associations in the horse.

P0337: Equine

Investigating Ancient Introgression between Caballine and Non-Caballine Equids

Kalpani de Silva, University of Louisville, Louisville, KY

The caballines and non-caballines are thought to have diverged approximately 4.5 million years ago. In a recent study, they have reported alleles for the gene *CXCL16* in the horse based non-synonymous variants at 4 locations in chr11 exon 1 producing two alternative proteins, *CXCL16*S and *CXCL16*R. The susceptible *CXCL16*S allele which is found to be fixed among other non-caballines acts as a receptor for Equine Arteritis Virus while *CXCL16*R lacks this ability. This type of Hybridization between species is understudied. Because our search is for events that happened so long ago, the haplotype blocks for which we are searching are likely much smaller than those produced in introgression events that have been reported in humans and Neanderthal necessitating a different search strategy.

This study is conducted to find whether *CXCL16*S allele entered the caballine lineage before or after several million years of divergent evolution. We are looking for other such events across the genome which might explain adaptive introgression in modern horses. Our study consists of NGS data from 6 animals including three Horses as well as three non-caballine equids, a Kiang, Zebra, and a Somali ass. The ancestral genome is inferred based on the fixed alleles among caballines and non-caballines. The genomes are scanned in small windows using Python and R packages to calculate D statistic which is specially designed to detect introgression from one population to another.

P0338: Equine

A Haplotype Map and Imputation Resource in the Horse

Robert J Schaefer, James R Mickelson and Molly E McCue, University of Minnesota, St. Paul, MN

The recently developed Equine genotyping array (MNEc670k) was designed for genotype imputation. Variants (SNPs) on the array were chosen to differentiate among patterns of linkage disequilibrium (LD) between diverse horse populations. This enables a smaller set of SNPs (~670k) to be imputed to a higher density (~2M) using haplotypes characterized in a reference population. Moreover, LD decays at different rates throughout the genome as well as between different breeds resulting in differing haplotype lengths. This is important for association studies where markers are unlikely to be functional, instead being in LD with the causal variant.

Using a reference population of 485 horses representing 15 breeds, we calculated breed specific haplotypes throughout the genome around each of 2M core SNPs. We cataloged these haplotypes along with LD decay in a relational database to enable data-driven SNP-to-gene mapping for association studies. Using this same reference population, we've implemented an equine imputation resource. Input samples genotyped on the MNEc670k array (or previous generation arrays) are phased and phylogenetically clustered in order to determine an imputation population. Here we present expected imputation accuracies to 2M markers and above for different breeds and from different original SNP densities.

P0339: Equine

Equine SNP Parentage Testing - Making the Transition from Microsatellites

Paul Flynn, Weatherbys Scientific, Naas, Ireland

Transitioning from Microsatellites to SNP molecular markers to perform parentage testing is becoming an ever more discussed topic within the global Equine community. A key area that requires exploration is ensuring maintenance of test integrity when moving to SNP based parentage. To investigate this area, Weatherbys Scientific has undertaken a pilot study to 1) review feasibility of performing Equine parentage testing using Blood and Hair samples as starting material to establish suitable SNP genotype profiles and 2) review the accuracy of performing Equine parentage testing using SNP genotype profiles in comparison to Microsatellite genotype profiles.

The study consisted of n = 186 animals across 9 Equine groups with sample sources including Blood and Hair. All animals were SNP genotyped using the Affymetrix Equine 670K SNP Array. A range of previously Microsatellite based parentage analysis categories were represented within sample selection i.e. Full (Sire/Dam/Offspring), Partial (e.g. Sire/Offspring), One Marker Exclusion, Complete Exclusions and Search by Sire/Dam. SNP parentage testing was performed across all case categories using 106 recommended ISAG SNP markers and a supplementary Weatherbys 1K SNP panel.

All bloods and ~80% of hairs proved suitable for SNP genotype profiling. SNP parentage testing accuracy was comparable to using Microsatellite approach for Full (Sire/Dam/Offspring) test categories. SNP parentage testing accuracy for Partial (e.g. Sire/Offspring) test categories benefited from supplementary Weatherbys 1K SNP panel. This pilot study provides a greater understanding of sample performance using Blood and Hair and test accuracy when applying a SNP based method for Equine parentage testing.

P0340: Equine

Optimization of Equine ChIP-Seq for the Functional Annotation of Animal Genomes

Nicole B. Kingsley[1], Carrie J. Finno[1], Geoffrey Berguet[2], Jan-Jaap Wesselink[2], Mustafa Tammoh[2], Catherine Creppe[2], Sol Schvartzman[2], Céline Sabatel[2], Colin Kern[3], Erin N. Burns[1], Huaijun Zhou[3], Jessica L. Petersen[4] and Rebecca R. Bellone[5], (1)University of California, Davis, Davis, CA, (2)Diagenode, Liège, Belgium, (3)Animal Science, University of California, Davis, CA, (4)University of Nebraska, Lincoln, NE, (5)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

A major aim of the Functional Annotation of Animal Genomes (FAANG) initiative is to characterize tissue-specific genomic regulation in agricultural species, including the horse. To that end, we are using Chromatin Immuno-Precipitation followed by sequencing (ChIP-seq) to identify four histone marks (H3K4me1, H3K4me3, H3K27ac, and H3K27me3) in eight prioritized equine tissues (heart, lung, liver, skeletal muscle, adipose, ovary, laminae, and parietal cortex). As investigation of histone modifications with ChIP-seq has not been performed extensively in horses, we first aimed to optimize experimental procedures through quality control testing (QC). Liver and adipose were selected for initial optimization. To collect sufficient immuno-precipitated DNA for Illumina sequencing, various amounts of snap-frozen tissue and shearing times were tested on each tissue with two antibody concentrations for each mark. Optimal conditions utilized 30mg of tissue and 12 shearing cycles for liver and 850mg of pooled tissue and 3x10 cycles for adipose to obtain 380-750ng of DNA for immuno-precipitation. Two analysis pipelines with different peak-callers, Sicer and MACS2, were used to evaluate the data for quality based on ChIP-seq standards established by ENCODE. All marks met or exceeded the standards with the exception of H3K27me3 in liver, which requires further optimization. From the analyses, we also determined that 0.5ug of antibody was sufficient for all successful marks except H3K27ac, which showed an increased number of peaks in both tissues using 1.0ug. ChIP-seq experiments on the two optimized tissues harvested from two Thoroughbred mares and QC of the six additional prioritized tissues are currently underway.

P0341: Equine

Transitioning from Illumina to Affymetrix: Platform Concordance and Lessons Learned

Heather M. Holl and Samantha A. Brooks, University of Florida, Gainesville, FL

Reliable, high quality genotypes are vital for genomic studies. In the horse, two Illumina BeadChips, the Equine SNP50 and SNP70 arrays, have been successfully used in a variety of studies involving domestication, population genetics, genetic selection, and trait mapping. More recently, an Affymetrix 670k Axiom array was developed in order to increase marker coverage for genomic studies. However, concordance rates between legacy and whole genome sequencing genotypes have not yet been evaluated. We obtained Axiom 670k genotype data from 767 horses for use in multiple projects. DNA was extracted from multiple sample types, and arrays were genotyped by two providers in four batches. Initial quality control analyses indicated problematic batch effects across service labs, thus we evaluated more stringent filtering criteria. Genotypes from alternate platforms were available for a subset of horses: seventeen horses with Axiom 670k and Illumina SNP50, six horses with Axiom 670k and Illumina SNP70, and eight horses with Axiom 670k and Illumina HiSeq. Overall genotype concordance rates

were determined separately for each Affymetrix defined SNP cluster classification. Additionally, we called variants from Illumina HiSeq lanes from twenty-one horses using the GATK HaplotypeCaller pipeline and identified variants incorrectly annotated as bi-allelic SNPs in the Axiom 670k reference files. We will present recommendations for genotyping and quality control for Affymetrix based projects.
This study was made possible in part by NPRP Grant 6-1303-4-023 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

P0342: Equine

Update on the Equine FAANG Initiative: How the Community Is Using the Biobank

Erin N. Burns[1], Theodore S. Kalbfleisch[2], Jessica L. Petersen[3], Rebecca R. Bellone[4] and Carrie J. Finno[1], (1)University of California, Davis, Davis, CA, (2)University of Louisville, Louisville, KY, (3)University of Nebraska, Lincoln, NE, (4)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

The Functional Annotation of the Animal Genome (FAANG) project aims to identify functional regulatory elements in animal genomes in both sexes and across multiple stages of development. For the horse, a biobank of tissues and cells was created, which includes 80 tissues and three cell types from two adult Thoroughbred mares. Last year, we reported on the creation of this biobank to be used by the equine community in the functional annotation of the genome in accordance with FAANG Consortium guidelines. Since the Plant & Animal Genome conference in January of 2017, the FAANG biobank was used in the generation of 7 different datasets. In collaboration with the equine genetics community, reduced representation bisulfite sequencing, microbiome sequencing, and chromatin immunoprecipitation sequencing have been conducted. In addition, to date, RNA has been isolated from 34 tissues of both biobank mares using Trizol® chloroform phase separation and clean up via Zymo Research columns. Extracted RNA had an average RIN score of 8.2 prior to stranded library preparation. mRNA and smRNA were sequenced on the Illumina HiSeq 4000. Data are publicly available and linked to the phenotype information within the FAANG database. Analysis of the RNA-seq data is ongoing to determine specific signatures of each tissue type.

P0343: Equine

Refined Phenotypes Associated with the Equine FAANG Biobank: Microbiome Sequencing

Shauna M. Tietze[1], Erin N. Burns[2], Rebecca R. Bellone[3], Carrie J. Finno[3] and Jessica L. Petersen[1], (1)University of Nebraska, Lincoln, NE, (2)University of California, Davis, Davis, CA, (3)University of California - Davis, Davis, CA

The FAANG (Functional Annotation of Animal Genomes) Project seeks to create a comprehensive database of functional elements in animal genomes. The utility of such a database is enhanced when specific phenotypic information is recorded on the animals studied. To that end, this work was performed to characterize the gastrointestinal (cecum, stomach, large and small colon) microbiome diversity of the two FAANG Thoroughbred mares. The V4 region of the 16S rRNA gene was amplified from isolated DNA of each sample using 515F and 806R primers. The amplified product was normalized, pooled, and sequenced on an Illumina MiSeq. Operation Taxonomic Units (OTUs) were classified to the species level and further subsampling allowed for composition, phylogenetic relationships, abundance, and diversity of the microbial taxon to be characterized. A high proportion of taxa present were sampled as denoted by a Good's Coverage Index averaging 92%. The colon samples had the highest microbial diversity, species richness, and OTUs, followed by the cecum, intestine, and stomach, respectively. Microbial composition differed across the regions analyzed. The stomach consisted of 96% Firmicutes, which comprised 30-55% of the other regions, in which Bacteroidetes were abundant. Verrucomicrobia and Proteobacteria were identified in the cecum, intestine, large colon, small colon, and stomach. While diversity, abundance, richness, and microbial composition differ among regions, diversity within each region was similar between the two horses. These data, along with genome regulation data from corresponding tissues in the same horses, will be useful for future comparison across and within species.

P0344: Equine

Identification of Imprinting in *Rtl1/RTl1*_Antisense using High-Throughput Analysis of RNA and DNA

Pouya Dini[1,2], Shavahn C. Loux[1], Alejandro Esteller-Vico[1], Kirsten E. Scoggin[1], Theodore S. Kalbfleisch[3], Peter Daels[2] and Barry A. Ball[1], (1)Department of Veterinary Science, University of Kentucky, Lexington, KY, (2)Department of Reproduction, Faculty of Veterinary Medicine, University of Gent, Merelbeke, Belgium, (3)University of Louisville, Louisville, KY

Chorioallantoic membranes (CAM) of the placenta provide an essential feto-maternal interface. One of the pivotal genes in the development of the CAM is the retrotransposon Gag like 1 (*Rtl1*) gene. In mice, *Rtl1* is paternally expressed, imprinted gene with an overlapping maternally expressed antisense transcript (*Rtl1_anti*), located at chromosome 12q32. *Rtl1_anti* contains several microRNAs including miR-127, and miR-136, among others. We have demonstrated the expression of *Rtl1/RTl1_anti* in equine CAM (unpublished) but until now there is no information available on the imprinting status of *Rtl1/RTl1_anti* in the equine genome. We applied stranded, paired-end RNA sequencing (RNA-Seq) to CAM tissue (N=6) and genomic sequencing (DNA-Seq) to sires and dams from reciprocal offspring. Sequencing data were trimmed by TrimGalore and aligned to the horse genome by STAR (V 2.5.2b) for RNA-Seq and by the Burroughs-Wheeler Aligner (V 0.7.12) for DNA-Seq. Subsequently, RNA-Seq reads were divided according to the strand of origin (positive and negative strands). Expressed alleles at *Rtl1/Rtl1_anti* loci in CAM were compared relative to the respective sire and dam genome. Parental status of alleles were identified in 6 trios (sire, dam and offspring), using Genome Analysis Toolkit (GATK). There were 46 informative SNPs at the investigated loci. Exact binomial test showed significant paternal expression in positive strands (93%) and significant maternal expression in negative strands (89%) at *Rtl1/Rtl1_anti* loci and, parental-specific expression was confirmed by a Chi-Square ($P<0.0001$). These data confirm that *Rtl1 and Rtl1_anti* (miR-127, miR-136, miR-433, miR-431 and miR-432) are reciprocally imprinted in the horse.

P0345: Equine

Annotation of a Structural Polymorphism in the Lath Gene Region of the Equine and Related Species

Katelyn M. Palermo, Kelsey Yates, Laura Patterson Rosa, Heather M. Holl and Samantha A. Brooks, University of Florida, Gainesville, FL

Cooling mechanisms are vital to thermoregulation and survival of all mammals. Equids and higher primate species have uniquely evolved to produce sweat as a primary means of cooling. Unlike humans, equine sweat is high in protein, specifically Latherin, a surfactant protein which increases pelt wetting and aids in evaporation. Encoded by the gene *LATH*, the locus is a paralog to human Bactericidal Permeability-Increasing Protein Family A Member 4 (*BPIFA4*) and representative of the rapidly evolving PLUNC (palate, lung and nasal epithelium clone) protein family. A previous study observed a polymorphic copy number variant (CNV) encompassing the *LATH* gene region of the domesticated horse and varying in copy number among horses of diverse breeds. In the horse, this particular structural polymorphism is under positive selection, possibly as a result of adaptive evolution for survival, or due to human-selection throughout domestication. Here we will report the results of quantifying exact copy numbers of five BPI genes in multiple equid species and across the Perissodactyl family which includes the rhinoceros and tapir. Our use of the QX200™ Digital Droplet™ qPCR and Illumina's Eco Real-Time PCR has allowed us to generate Cq numbers that illustrate a large difference between gene expression of *LATH* in horses (25.73446529) versus that of rhinos (30.15830968). Future work will include testing such individuals using the Qualitative Intradermal Terbutaline Sweat Test (QITSTs) and quantification of the amount of *LATH* protein produced in sweat samples by western blot.

P0346: Equine

Epigenomic Diversity in the Mammalian Brain

Chongyuan Luo, Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA

The epigenome is an ensemble of chemical modifications of DNA and chromatin that modulates the activities of the genome, which plays instrumental roles in gene regulations in healthy and disease tissues. We have developed a sequencing based method to profile DNA methylation (mC) at single-base resolution across the whole mammalian genome. Using this approach, we found high levels of non-CG methylation (mCH) at locations throughout the genome in human and mouse brains. Mammalian brains contain numerous types of neurons that can be distinguished by their morphological, physiological and functional properties. Using an epigenomics dataset produced from nuclei of specific neuronal types purified by INTACT approach, we identified abundant epigenomic and gene-expression differences across three excitatory and inhibitory neuron types in adult mouse. Extending cell type specific mC analysis to all brain cell types requires unbiased single cell mC profiling. We developed a new method for high-throughput single nucleus methylcytosine sequencing (snmC-seq). Using >6,000 single cell methylomes, we identified 16 mouse and 21 human neuron types in the frontal cortex. The results suggest expanded neuronal diversity in the human cortex, which is consistent with the finding of a human-specific inhibitory neuron subpopulation. Our epigenomic analysis allowed the prediction of approximately ~500,000 cell type specific regulatory elements for mouse or human neuron types. Currently we are generating single cell methylomes for cell type classification of the whole mouse brain. Our single nucleus methylome approach can be applied to all human tissues for producing epigenomic profiles to inform the human cell atlas.

P0347: Equine

Who Watches the Watchers? Profiling miRNA Expression along the Equine Gastrointestinal Tract

Kailee Reed and Stephen J. Coleman, Colorado State University, Fort Collins, CO

Maintaining homeostasis is critical to proper gastrointestinal function. Dynamic interactions between the microbial population and its equine host are at the center of this equilibrium. Non-coding RNAs, specifically microRNAs (miRNA), have recently been implicated as a key mechanism for regulating aspects of the interaction between host and microbiota. miRNAs are single-stranded 20-22 basepair molecules and play an active role in post-transcriptional regulation of protein-coding transcripts. Few experiments have investigated miRNA expression in equine gastrointestinal tissues or generated a profile of their expression along the tract. This pilot study is intended to address this deficiency by characterizing miRNA expression patterns in tissues of the equine hindgut. Total RNA was isolated using TRIzol^{TM} from intestinal epithelium collected at the cecum (base and apex), right and left ventral colon, pelvic flexure, right and left dorsal colon, and the small colon (n=3 for each tissue). miRNA was reverse-transcribed using the miScript II RT kit. Relative transcript abundance was quantified using the miScript SYBR green PCR kit with primer sets designed for 346 annotated mature equine miRNAs. Biological replicates were pooled and differential expression of miRNA transcripts was determined following normalization by the delta $C_t = C_t(\text{target}) - C_t(\text{mean of sample})$ method. Profiling miRNA expression throughout the equine hindgut will help improve understanding of the regulatory processes involved in gastrointestinal homeostasis.

P0348: Equine

Unintended Consequences: Characterizing the Impact of Combined Moxidectin and Praziquantel Anthelmintic Treatment on the Equine Microbiome

Isabelle Kunz and Stephen J. Coleman, Colorado State University, Fort Collins, CO

The gastrointestinal microbiome (GIM) plays an essential role in maintaining intestinal homeostasis. Disruptions to the GIM can have profound effects on the wellbeing of the host organism. Infection with large strongyles parasites is a significant issue in horses. Correspondingly, the use of anthelmintic drugs for parasite control is standard practice in the equine industry. To date, few studies have investigated the impact of anthelmintic treatment on the horse GIM. This pilot study evaluates the hypothesis that anthelmintic drug treatment will alter the fecal microbiome. Twelve horses were administered a single dose of QUEST® PLUS [active ingredients: Moxidectin and Praziquantel] (Zoetis US) on day 0; fecal samples were collected on day -3 (pre-treatment), day 2 (post-treatment 1), day 9 (post-treatment 2). DNA was isolated from the samples and the V4 region of the bacterial 16S rRNA gene was amplified for Next Generation Sequencing, generating an average of 147,645 +/- 35,496 reads per sample. Amplicon sequencing data was quality filtered, processed and analyzed using the Quantitative Insights into Microbial Ecology (Qiime 1.9) software. Differential abundance of operational taxonomic units (OTUs) between day –3 and day 2 was determined using the DESeq2 algorithm. Forty-six OTUs had significantly different abundance (padj-value<0.05) immediately following treatment, the majority (34/46) of which were decreased abundances. Twenty-four (52%) of the differentially abundant OTUs represented the order *Clostridiales*, which have been associated with fiber digestion and immunity. These results suggest that combined administration of Moxidectin and Praziquantel alters the equine GIM and may have functional consequences for the host.

P0349: Equine

Utilization of the Foal Fecal Microbiota to Understand Gut Flora Transitions: From Birth to Weaning

Michael J. Mienaltowski[1], Ubaldo De La Torre[2], John D. Henderson[2], Kathleen Furtado[2], Madeleine Leahy[2] and Elizabeth A. Maga[2], (1)University of California, Davis, Davis, CA, (2)University of California Davis, Davis, CA

Few studies have been conducted to understand the gut microbiota of foals. A healthy gastrointestinal (GI) tract with a properly established microbiota is necessary for a foal to develop into a healthy weanling. We hypothesized that the establishment of the gut flora in foals is directly correlated to diet and environmental exposures, and could be assessed from analyses of the fecal microbiota during GI transition. Fecal samples from 42 sets of foals and mares were collected at multiple time points ranging from birth to weaning. Bacterial DNA was isolated and the V4 domain of bacterial 16S rRNA genes was amplified and then applied to next-generation sequencing to characterize the fecal microbiota in each foal. Microbial taxonomic assignment and relative abundance determination were performed using QIIME (Quantitative Insights Into Microbial Ecology). Specific comparisons in microbial populations were made using LefSe (LDA Effect Size). STAMP (STatistical Analysis of Metagenomic Profiles) was used to characterize functional roles of microbial populations in host biology. We found that bacterial population compositions followed a pattern throughout the early life of the foal in an age-dependent manner. Moreover, we were also able to recognize differences in microbial populations amongst diarrheic foals. Future efforts will better discern the effects of lesser abundant bacterial populations that may be just as essential to GI health, as well as characterize microbial populations at additional time points to further investigate the fecal microbiota as the foal transitions to weaning.

P0350: Equine

Gene Loci and SNP Haplotypes from the Horse NK Receptor Regions

Donald Miller[1], Theodore S. Kalbfleisch[2], James N. MacLeod[3], Ernest Bailey[4], Heather M. Holl[5], Samantha A. Brooks[5], Joel A. Malek[6] and **Douglas Antczak**[1], (1)Cornell University, Ithaca, NY, (2)University of Louisville, Louisville, KY, (3)Department of Veterinary Science, University of Kentucky, Lexington, KY, (4)University of Kentucky, Lexington, KY, (5)University of Florida, Gainesville, FL, (6)Weill Cornell Medical College, Doha, Qatar

The equine genome contains two Natural Killer Receptor (NKR) regions, with lectin type LY49-like genes covering 600KB on Eca6, and a Killer-cell Immunoglobulin-like Receptor (KIR) complex spanning 1.25MB on Eca10. Earlier work showed the horse LY49 region expanded from a single gene consistent with other mammals like cattle and swine, to now include six functional LY49-like genes (Takahashi et al., 2004). The KIR locus has multiple copies of both LILRA and LILRB (Leukocyte Immunoglobulin-like Receptor subfamily alpha and beta) genes, although the expression pattern of each gene is yet to be determined and previous research has indicated that many of these loci may have lost function (Takahashi et al., 2004). To examine gene structure in both regions, we used a *de novo* assembly of 10x Genomics™ Chromium™ linked reads from Twilight, the DNA donor of the horse whole genome sequence, and compared this to the EquCab2 reference assembly and annotation. To study sequence variation in the NKR regions, we examined SNP haplotypes from horses tested on the Affymetrix 670K horse SNP chip. From a sample set of 26 horses, we identified five well defined SNP haplotypes for each region. These data expand our knowledge of the two NKR complexes in the horse, and allow us to more clearly define the expressed genes and sequence variation found therein. This study was made possible in part by NPRP Grant 6-1303-4-023 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

P0351: Equine

Sex Determination of Equine Embryos Using the Ratio of Heterozygous to Homozygous Variants and Differential Gene Expression of Chromosome X

Alejandro Esteller-Vico[1], Pouya Dini[1], Theodore S. Kalbfleisch[2] and Barry A. Ball[1], (1)Department of Veterinary Science, University of Kentucky, Lexington, KY, (2)University of Louisville, Louisville, KY

The presence of two Chromosome X (ChrX) in females results in differences in the DNA content and on sexually dimorphic gene expression that can be detected during early stages of development in mammals. The objectives of this study were to determine the sex of equine embryos using RNAseq data of ChrX from embryos obtained between days 12 and 16 (n=15). Because males are hemizygous for ChrX, variant discovery and genotyping software should identify a minimal number of heterozygosities in males and localized in the pseudo autosomal region of ChrX. Data was generated on a HiSeq2000 generating an average of $12.3x10^6$ stranded 150-bp paired-end reads per sample. To determine the sex of the embryos, the ratio of heterozygous variants to homozygous variants was calculated on ChrX for 15 equine embryos and 15 mare endometrial samples as controls using Unified Genotyper software. The results for the ratio of heterozygous variants to homozygous variants (heterozygous variants/(heterozygous variants+homozygous variants)) segregated the embryos in two tight groups. The ratio was 0.199±0.011 in group one and 0.342±0.01 in group two. The ratio in group two was similar to the ratio of mare endometrial samples (0.365±0.027) and both were significantly higher than group one (p-value<0.0001). Therefore, group one was identified as male embryos and group two as female embryos. The sex of the embryos in each group was confirmed with sex specific transcripts. However, the ratio in males was higher than expected and these heterozygosities were not localized in the pseudo-autosomal region of ChrX.

P0352: Equine

Characterization of Pregnancy-Associated miRNA Clusters in the Horse

Shavahn C. Loux[1], Pouya Dini[1], Kai Wang[2], David Baxter[2] and **Barry A. Ball**[1], (1)Department of Veterinary Science, University of Kentucky, Lexington, KY, (2)Institute for Systems Biology, Seattle, WA

Initial transcription of microRNAs (miRNA) often results in long, primary miRNAs which include multiple mature miRNA. These simultaneously transcribed miRNAs are known as clusters and often work synergistically to influence physiological process such immunity or pregnancy. Several pregnancy-associated clusters have been described, including the C19MC cluster in primates and the C14MC cluster in eutherian species; we aimed to identify additional pregnancy-related clusters present in the mare. We used next-generation sequencing to characterize the small RNA transcriptome in the placenta (chorioallantois) and serum of mares at 4 m, 10 m gestation and post-partum (n = 3 / stage), as well as the serum of diestrous mares (n = 6). Using these data, we identified all annotated miRNAs which were < 10 kb apart based on stated genome location via miRbase (v. 21). In total, sixty-one clusters were identified, with the majority located on ECAX (11 clusters, 61

miRNA), ECA24 (2 clusters, 46 miRNA), and ECA11 (6 clusters, 12 miRNA). To classify a cluster as pregnancy-associated, a minimum of two miRNA and 50% of the miRNA within the cluster had to be differentially regulated during pregnancy ($P < 0.01$). Overall, 14 out of 61 total clusters were identified as pregnancy-associated, including 108 associated miRNAs (64 differentially expressed; 42 in placenta, 35 in serum, 13 in both). These clusters included the previously characterized human C14MC, as well as immune-related clusters such as miR-181a/181b and miR-221/222. The identification of novel, pregnancy-associated miRNA clusters should help elucidate the role of miRNAs during equine pregnancy.

P0353: Equine

Determining Copy Numbers of Key Fertility Genes in the Equine Y Chromosome

Caitlin Castaneda, Texas A&M University, College Station, TX; Duquesne University, Pittsburgh, PA

The mammalian Y chromosome houses essential male fertility genes. Previous mammalian studies provide evidence that copy number variations (CNVs) in these genes affect male reproductive phenotypes. Here, we initiate CNV analysis of multicopy genes in the horse male specific Y (eMSY) using digital droplet-PCR. Our goal is to determine a normal range of CNVs for all multicopy genes (total 15, of which 8 are novel and horse specific) in a cohort of normal male horses of diverse breeds. The data will form a baseline for the discovery of variants associated with reduced stallion fertility. Initial studies involved 64 normal males of seven breeds, and three infertile stallions. The breeds studied were Thoroughbred (including the Thoroughbred stallion *Bravo* - the DNA donor of the eMSY reference sequence), American Quarter Horse, Arabian, Suffolk, Caspian Pony, Mongolian Native, Icelandic, Shetland Pony, Mongolian and Lipizzaner. TaqMan assays were designed for five multicopy genes: *TSPY*, *RBMY*, *UBA1Y*, and *HSFY*, and for the *SRY* gene. Using autosomal myostatin (*MSTN*) as the single copy reference gene, multiple ddPCR iterations provided consistent CNs amongst all seven breeds and confirmed that *SRY* is a single copy gene. None of the studied genes showed different CNs in infertile stallions. However, for all genes we observed less copies by ddPCR compared to the Y reference assembly. Future studies include analysis of the remaining multicopy genes, including equine specific transcripts such at *ETSTY2*. Gaining information on eMSY CNVs would enable improving the assembly of eMSY multicopy regions, and can potentially lead to the discovery of CNVs affecting stallion fertility.

P0354: Equine

Genome Wide Association Study Identifies Locus for the Mushroom Coat Color Dilution in Shetland Ponies

Jocelyn L. Tanaka, University of California - Davis, Davis, CA, Maura Mack, Colorado State University Veterinary School, Fort Collins, CO, Tosso Leeb, University of Bern - Institute of Genetics, Bern, Switzerland, Christine Flury, Bern Univeristy of Applied Sciences, Department of Agriculture Forestry, Food Science & Management, Bern, Switzerland, Iris Bachmann, University of Bern, BERN, Switzerland, M. Cecilia Penedo, Veterinary Genetics Laboratory, University of California, Davis, Davis, CA and Rebecca R. Bellone, Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Coat color is a trait of economic importance to horse breeders. Mushroom is a unique phenotype in the Shetland pony that is characterized by the dilution of the chestnut coat color to a light sepia tone, while leaving bay and black base coat colors unaffected. The molecular mechanism for this trait is unknown. Previous pedigree analysis suggested a simple recessive mode of inheritance. To identify a candidate locus for further investigation a genome wide association study (GWAS) utilizing the Affymetrix 670K array (MNEc670k) was performed with DNA isolated from 12 mushroom and 12 chestnut horses. To correct for population substructure, (λ=1.11), a single locus mixed linear model analysis (EMMAX) approach was utilized. This approach identified a single region on ECA7 that reached genome wide significance ($P_{corrected}=7.64X10^{-5}$). This region contains a 328 Kb haplotype that was perfectly concordant with the mushroom phenotype. Replication testing of four SNPs from this haplotype using 42 additional horses confirmed this association ($P_{combined}=4.51X10^{-16}$). The haplotype spans eight genes, one of which is a putative functional candidate. Further research is needed to explore this candidate's role in phaeomelanogenesis and mushroom coat dilution.

P0355: Equine

Selective Sweep Mapping Using a Unique Nordic Horse Model Revealed *EDN3* as a Candidate Gene for Harness Racing Performance

Kim J. Fegraeus[1], Brandon D. Velie[1], Jeanette Axelsson[1], Rachel Ang[2], Natasha A Hamilton[2], Leif Andersson[3], Jennifer R.S. Meadows[3] and **Gabriella Lindgren**[1], (1)Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, (2)School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia, (3)Dept of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden

The unique origin of the Swedish-Norwegian Coldblooded trotter makes the breed ideal for studying the genetics of racing performance. By comparing the genomes of Coldblooded trotters, Standardbreds and North-Swedish draught horses for a large number of single nucleotide polymorphisms (SNPs), the aim of the study was to identify genetic regions that may be under selection for racing performance. A fixation index (Fst) analysis was performed and sliding window Delta Fst values were calculated across the breeds using data generated from the equine SNP50K array (Coldblooded trotters, n=11; North-Swedish draught horses, n=19; Standardbreds, n=12). The average Delta Fst was calculated for windows of five SNPs and the five top windows, where the Fst between Coldblooded trotters and Standardbreds was low and the Fst between Coldblooded trotters and North-Swedish draught horses was high, were chosen for further investigation. Associations between the top SNP from each region and harness racing performance was analyzed in 400 raced Coldblooded trotters. One SNP (g.22:45748491C>T) showed significant associations with racing performance, with the CC genotype appearing to be negatively associated with the majority of performance traits tested. In addition, four consecutive SNPs spanning 1.5 kilobases (kb) all showed significant associations with harness racing performance. Further, the SNP identified was genotyped in 1,634 horses of 14 different breeds. The frequency of the TT genotype was high in breeds typically used for racing and show jumping while the frequency of the CC genotype was high in most pony breeds and draught horses. The closest gene in the top region identified was the *Endothelin3* gene (*EDN3*).

P0356: Equine

Investigation of Candidate Genes Involved in Dysregulated Vitamin E Metabolism in Equine Neuroaxonal Dystrophy(eNAD)

Sichong Peng[1], Erin N. Burns[1], Matthew Bordbari[2], Stephanie J. Valberg[3] and Carrie J. Finno[2], (1)University of California, Davis, Davis, CA, (2)University of California - Davis, Davis, CA, (3)Michigan State University, East Lansing, MI

Background. Equine neuroaxonal dystrophy (eNAD) is an inherited neurodegenerative disorder in horses. Previous clinical studies show evidence that a vitamin E deficiency is likely to be involved in the development of eNAD.
Objective. We aimed to find genetic variations that are associated with eNAD in several candidate genes responsible for vitamin E metabolism.
Animals. A total of 62 client owned eNAD-affected horses and 82 client owned control horses were used in this study.
Methods. Candidate Gene Association study (CGAS) was performed on horses whose whole genome sequence (WGS) dataset was previously obtained. Multiplex genotyping was then performed on prioritized variants on an additional population for confirmation purpose. Fisher's exact test was used in case-control study to determine significance of association between variants and the eNAD phenotype.
Results. One variant at chr17:71585141 was identified to be significantly associated to eNAD phenotype ($P_{Bonferroni}$=0.001887). However, this variant was found at similar allele frequency in eNAD affected population as in general population in SRA database.
Conclusions and clinical importance. Chr17:71585141 (C>T) is a potential variant for further investigation. Currently the disease can only be diagnosed postmortem. Discovery of associated variants may lead to development of genetic diagnosing tools which can greatly aid early clinical diagnosis and treatment as well as shed light on the etiology of the disease as it remains unclear.

P0357: Equine

Proteome and Transcriptome Profiling of Equine Myofibrillar Myopathy Identifies Diminished Peroxiredoxin 6 and Enhanced Cysteine Metabolic Pathways

Sudeep Perumbakkam[1], Stephanie J. Valberg[1], Erica C. McKenzie[2] and Carrie J. Finno[3], (1)Michigan State University, East Lansing, MI, (2)Oregon State University, Corvallis, OR, (3)University of California-Davis, Davis, CA

Equine myofibrillar myopathy (MFM) causes exertional muscle pain and is characterized by myofibrillar disarray and ectopic protein aggregates of unknown origin. To investigate the pathophysiology of MFM, we compared the skeletal muscle proteome and 3 h post-exercise transcriptome of gluteal muscle in MFM and control Arabian horses using iTRAQ and RNA-sequencing analyses. Differential expression (DE) was evaluated using edgeR and pathway analysis using Cytoscape and Cluego. Proteome analysis revealed significantly lower antioxidant peroxiredoxin 6 content (PRDX6, ↓4.14 $\log_2$ fold change [FC]), sarcomere protein tropomyosin (TPM2, ↓3.24X) and higher fatty acid transport enzyme carnitine palmitoyl transferase (CPT1B, ↑3.49X) in MFM vs. control muscle at rest. Three hours after exercise, 191 genes were DE in MFM vs. control muscle with a remarkably focused > 1.5 $\log_2$FC in genes involved in sulfur compound/ cysteine metabolism such as cystathionine-beta-synthase [CBS, ↑4.51] and a cysteine and neutral amino acid membrane transporter [SLC7A10, ↑1.79]. In MFM vs. control at rest, 284 genes were DE with > 1.5 $\log_2$ FC in pathways for structure morphogenesis, fiber organization, tissue development and cell differentiation including> 2 $\log_2$ FC in alpha actin-cardiac [↑ ACTC1], cytoskeletal desmoplakin [↑ DSP], basement membrane usherin [↓ USH2A] and delta like non-canonical Notch ligand 1, [↓ DLK1]. In conclusion, myofibrillar disarray and protein aggregation in MFM horses was embodied by DE expression in pathways of structure/fiber organization and tissue regeneration. Reduced antioxidant capacity as a potential etiology for MFM was supported by diminished cysteine rich antioxidant peroxiredoxin 6 with compensatory increased cysteine synthesis following exercise.

P0358: Equine

Additional Evidence Supports DDB2 T338M As the Genetic Risk Factor for Ocular Squamous Cell Carcinoma in Horses

Moriel Hana Singer-Berk[1], Mary Lassaline[2], Kelly Knickelbein[3], Nikhil Joshi[4], Monica T. Britton[4], Matthew Settles[4], Zachary T. Lounsberry[1], Christopher Reilly[5], Ellison Bentley[6], Alison Clode[7], Cord Drögemüller[8], Lucia Unger[9], Vincent Gerber[9], Tammy Miller-Michau[10] and Rebecca R. Bellone[1], (1)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA, (2)UC Davis School of Veterinary Medicine, Davis, CA, (3)Veterinary Medical Teaching Hospital, University of California-Davis, Davis, CA, (4)UC Davis Genome Center, Davis, CA, (5)Insight Veterinary Specialty Pathology, Austin, TX, (6)University of Wisconsin-Madison, School of Veterinary Medicine, Madison, WI, (7)Port City Veterinary Referral Hospital, Portsmouth, NH, (8)Institute for Genetics University of Bern, Bern, Switzerland, (9)Swiss Institute of Equine Medicine, University of Bern and Agroscope, Bern, Switzerland, (10)Bluepearl Veterinary Partners, Tampa, FL

Squamous cell carcinoma (SCC) is the most common periocular cancer, and the second most common tumor in horses. A missense mutation in *damage specific DNA binding protein 2 (DDB2, p.Thr338Met),* within an associated 483 kb haplotype on ECA12, was identified as a recessive genetic risk factor for limbal SCC in Haflinger horses. To determine if this variant also contributes to risk for SCC in other ocular and urogenital locations and in other breeds commonly diagnosed with SCC, it was genotyped in Haflingers (N=110), Belgians (N=36), Appaloosas (N=42), and Arabs (N=25). These horses were diagnosed with SCC or were classified as unaffected based on clinical examination. A significant association was detected in Haflingers and Belgians when considering all ocular locations ($P= 2.80 \times 10^{-16}$ and $P= 1.98 \times 10^{-5}$ respectively). This association was not perfectly concordant with phenotype, as 24% of the Haflingers and Belgians affected with ocular SCC were not explained by homozygosity for this mutation. Therefore, high throughput sequencing data from six Haflingers (four cases and two controls) were analyzed to identify additional variants for investigation. Sixty-seven polymorphisms from the previously associated locus on ECA12 were genotyped in 103 Haflingers. Analysis revealed that no other variant from this locus explained the genetic risk better than *DDB2 p.Thr338Met* ($P = 7.83 \times 10^{-15}$). These data provide further support that the *DDB2* variant is a contributing risk factor for ocular SCC in Haflingers and Belgians and can be utilized as a diagnostic tool to inform clinical and breeding decisions.

P0359: Equine

Genome-Wide Association Study Identifies a Putative Risk Locus on ECA19 for Bilateral Corneal Stromal Loss in Friesian Horses

Kelly E Knickelbein[1,2], Mary Lassaline[3], Coral Alberi[1], Alyssa Atilano[1], Riccardo Stoppini[4], Hanneke Hermans[5], Willem Back[5,6], Michael J. Mienaltowski[7] and Rebecca R. Bellone[1,8], (1)UC Davis Veterinary Genetics Laboratory, Davis, CA, (2)UC Davis Veterinary Medical Teaching Hospital, Daivs, CA, (3)UC Davis School of Veterinary Medicine, Davis, CA, (4)Consultant in Equine Ophthalmology, Calvagese della Riviera, Italy, (5)Utrecht University, Utrecht, Netherlands, (6)Ghent University, Merelbeke, Belgium, (7)University of California, Davis, Davis, CA, (8)Veterinary Genetics Laboratory, University of California - Davis, Davis, CA

Bilateral Corneal Stromal Loss (BCSL), a condition recently described in Friesian horses, is characterized by bilateral progressive corneal stromal thinning in the inferior cornea without evidence of infection or malacia and with or without concurrent epithelial loss. Surgical stabilization of affected corneas can prevent progression to corneal perforation, irreversible ocular damage, and blindness. An X-linked recessive inheritance pattern was investigated based on the sex distribution of initial cases (8/9 male). Collection of additional cases supported an X-linked hypothesis but pedigree data were not conclusive (11 males, 4 females). Sequencing the X-linked candidate gene *biglycan*, a collagen-regulating extracellular matrix protein with known function in the cornea, revealed a SNP in the 3'UTR region that was fixed in Friesians but did not identify any variants segregating with the disease phenotype. Therefore, to identify additional loci for investigation and test an X-linked hypothesis further, a case-control genome-wide association study utilizing 10 BCSL-affected Friesians and 35 unaffected Friesians was performed. Basic allelic association revealed genome-wide significance ($P < 1.60 \times 10^{-7}$) for loci on ECA18, ECA19, and ECAX, although genomic inflation was high ($\lambda = 1.33$). Correcting for population substructure and the gender imbalance in our data set using a mixed linear model analysis supported a 539 Kb locus on ECA19 as significantly associated ($P = 3.76 \times 10^{-10}$) with this disease. This locus contains one annotated gene which is being investigated as a potential genetic risk factor for Bilateral Corneal Stromal Loss in Friesians horses.

P0360: Equine

RNA Sequencing Reveals HCN4 as a New Mediator of Airway Hyper-Responsiveness in a Spontaneous, Naturally Occurring Equine Asthma Model

Courtney Hunter[1], Lais Costa[2], Jim Cooley[3], Lauren Bright[4], Phillipe Chouvarine[5], Andrew Claude[6], Alison Eddy[7], Bindu Nanduri[7], Jacquelyn Bowser[8], Wellesley Dittmar[7], Alison Linch[7], Fiona McCarthy[9], Shane Burgess[9] and Cyprianna Swiderski[10], (1)Mississippi State College of Veterinary Medicine, Starkville, MS, (2)Mississippi State College of Veterinary Medicine, Mississippi State, MS, (3)Mississippi State University, Missisippi State, MS, (4)Rutgers University, New Brunswick, NJ, (5)Hannover Medical School, Hannover, Germany, (6)Michigan State University, East Lansing, MI, (7)Mississippi State University, Mississippi State, MS, (8)Johnson and Wales, Providence, RI, (9)University of Arizona, Tucson, AZ, (10)Department of Clinical Sciences, Mississippi State, MS

Equine Pasture Asthma (EPA), an asthma-like disease affecting horses housed on pasture in hot, humid climates, demonstrates key diagnostic features of severe, adult asthma. These include reversible airway obstruction, neutrophilic airway inflammation, and airway hyper-responsiveness (AHR) of a magnitude that is diagnostic of severe asthma (≤1mg/ml methacholine). Decreasing AHR decreases asthma severity, making it a primary goal of asthma therapy. To address this goal, we employed RNA sequencing of lung tissue from 2 horses with EPA and 2 controls to identify novel signaling pathways that contribute to AHR in EPA horses. By manually filtering 1376 differentially expressed genes (DEGs) that were a) conserved in diseased horses during clinical exacerbation, b) not differentially expressed by season in controls, c) had raw read counts approaching 0 during disease remission, and d) have known roles in autonomic signaling and muscle physiology, we identified hyperpolarization activated cyclic nucleotide gated potassium channel 4 (HCN4) as an overexpressed target for potentiating AHR in EPA horses. Histochemical staining of lung samples from a separate EPA and control cohort confirmed that increased HCN4 localizes to airway smooth muscle (ASM) in lung samples collected during seasonal pasture asthma exacerbations. Antagonist responses in isolated equine bronchi confirm that HCN4 signaling contributes to the contractile responses of ASM in horses. Coupled with the physiologic role of HCN4 in raising resting muscle membrane potentials, our findings support that HCN4-mediated current has a role in AHR that characterizes EPA horses.

P0361: Equine

Elucidating the Etiology of Atypical Equine Thrombasthenia in Thoroughbreds

Anna R. Dahlgren, Fern Tablin and Carrie J. Finno, University of California - Davis, Davis, CA

Bleeding disorders are frequently seen in Thoroughbred racehorses and can negatively impact their performance. Atypical Equine Thrombasthenia (AET) was elucidated as a frequent cause of bleeding, affecting approximately one in every 150 Thoroughbreds. AET can be diagnosed using flow cytometry to determine the extent that platelets bind fibrinogen. Affected platelets bind 35% or less fibrinogen as compared to healthy platelets. Pedigrees of affected horses indicate that the disease is heritable, though the underlying disease etiology remains unknown. AET platelets have been thoroughly phenotyped and the thrombin signaling pathway identified as the likely site of a causative genetic variation. Six AET affected and twelve unaffected Thoroughbreds underwent whole genome sequencing (50x coverage) and segregating variants were identified using both candidate gene and whole genome approaches. Due to the aberrant signaling in the thrombin pathway, fourteen genes that code for the primary proteins within the pathway were investigated using the candidate gene approach. There was no variation in the candidate genes significantly associated with AET. However, the whole genome approach identified a region of interest with segregating variants on chromosome 24. By identifying the genetic mechanism of AET, breeders can select against the disease to improve Thoroughbred health and performance.

P0362: Equine

Development of a Predictive Algorithm for Recurrent Exertional Rhabdomyolysis in Thoroughbred and Standardbred Racehorses

Samantha K Beeson[1], James R Mickelson[1], Rachel Ang[2], Natasha A Hamilton[2], Matthew M Binns[3], Sarah Blott[4], Paul Caputo[5], Cajsa M Isgren[6], Alison Moore[7], Richard J Piercy[8], June E Swinburne[9], Mark Vaudin[10] and Molly E McCue[1], (1)University of

Minnesota, St. Paul, MN, (2)School of Life and Environmental Sciences, The University of Sydney, Sydney, Australia, (3)Equine Analysis Systems, LLC, Midway, KY, (4)University of Nottingham, Nottingham, United Kingdom, (5)Paul Caputo, DVM, Parkland, FL, (6)Norwegian University of Life Sciences, Olso, Norway, (7)Moore Equine Services, Ontario, ON, Canada, (8)The Royal Veterinary College, London, United Kingdom, (9)Animal Health Trust, Newmarket, United Kingdom, (10)Animal Health Trust, Kentford, United Kingdom
Exertional rhabdomyolysis (ER) is a painful condition characterized by rapid-onset muscle cell necrosis and subsequent leakage of intracellular contents into the systemic circulation in response to exercise. Approximately 5-10% of Thoroughbred (TB) and 6% of Standardbred (STDB) racehorses suffer from recurrent exertional rhabdomyolysis (RER). Heritability for RER ranges from 0.34-0.42% in TBs and 0.39-0.45% in STDBs. Genome-wide association studies (GWAS) have identified nine genomic regions of interest (ROIs) significantly associated with RER in TB (n=491) and STDB (n=476) horses: two in TBs, six in STDBs, and one shared locus. In addition to 297 positional candidate genes located in GWAS ROIs, 2,115 biologic candidate genes for ER were identified by literature search, keyword-based gene prioritization, and use of publicly available differential gene expression data (36-gene overlap). Variants were identified from whole genome sequences (WGS) from 10 RER cases and 10 controls in each breed and imputed from low-density equine SNP array data in the entire GWAS cohort. Random forest (RF) feature importance scores for WGS variants were used to select potentially predictive variants in stages using window sizes of 50kb, 1Mb, chromosome, and genome. Based on allelic differences between cases and controls, location relative to positional and biologic candidate genes, predicted functional effect, and RF importance scores, 32,781 variants were selected for NuGen Allegro targeted genotyping. Results from this analysis will be used to assess WGS imputation accuracy from low-density equine arrays and validate putative functional and predictive variants for RER susceptibility in TB and STDB racehorses.

P0363: Equine

Prevalence of the Mutation Conferring Susceptibility to Immune-Mediated Myositis in Seven Performance Subgroups of American Quarter Horses

Giuliana M. Gianino[1], Stephanie J. Valberg[2], Sudeep Perumbakkam[2], Marisa L. Henry[2], Keri Gardner[2], M. Cecilia Penedo[3] and Carrie J. Finno[1], (1)University of California - Davis, Davis, CA, (2)Michigan State University, East Lansing, MI, (3)Veterinary Genetics Laboratory, University of California, Davis, Davis, CA
Equine immune-mediated myositis (IMM) is a disease in the American Quarter Horse (QH) breed characterized by recurrent, rapid-onset muscle atrophy and lymphocytic infiltration of myofibers, often following an episode of respiratory infection, vaccination, or rhabdomyolysis. Recently, a genome-wide association study and whole genome sequencing identified a functional variant which was significantly associated with the IMM phenotype. The purpose of this study was to estimate the frequency of this variant suspected to confer susceptibility to IMM in seven elite performance subgroups of QHs. Selection of elite QHs for this study was based on records obtained of the most competition points (halter discipline) or money earned (all other disciplines) in the last three years by individual horses registered with the American Quarter Horse Association (AQHA). With permission from the AQHA, top-performing horses from the barrel racing (42), cutting (43), halter (50), racing (36), reining (35), Western pleasure (45), and working cow (36) disciplines were genotyped for the IMM variant using pyrosequencing and a basic variant frequency was calculated in each subgroup. Of the total 287 elite performance horses genotyped, 91% were wildtype, 9% were heterozygotes, and 1 individual was homozygous for the novel variant. Frequency of the IMM variant was highest in reining (0.114), working cow (0.097), and halter (0.080) horses, followed by cutting (0.047) and Western pleasure (0.011) horses. The IMM variant was not observed in horses from the barrel racing or racing disciplines in this survey.

P0364: Equine

Chronic Idiopathic Anhidrosis in the American Quarter Horse - Don't Sweat over It.

Laura Patterson Rosa[1], Heather M. Holl[1], Neely L. Walker[2], Martha F. Mallicote[1], Robert J. Mackay[1] and Samantha A. Brooks[1], (1)University of Florida, Gainesville, FL, (2)Louisiana State University, Baton Rouge, LA
Chronic Idiopathic Anhidrosis (CIA) in horses is a condition characterized by a persistent reduction or complete lack of sweat. In order to dissipate excess body heat, the horse relies predominantly on evaporation of sweat from the skin surface. Horses in tropical climates are at higher risk, yet CIA is now frequently recognized in many non-tropical regions. Clinical signs may include labored breathing, hyperthermia, reduced appetite, decreased water intake, hair loss, dull hair coat and depression. CIA cases necessitate intensive management, restricted physical activity and often require retirement from breeding or competition. With no known cure and no effective medical treatment, anhidrotic horses can suffer severe consequences, including multiple organ failure in response to hyperthermia and, in some instances, death. An epidemiological study conducted in Florida revealed that diagnosis of CIA was associated with breed, use, birth place, and most importantly, family history. The odds are 21.67 times higher in horses with a family history of CIA, highlighting the strong genetic component to this disease. Our research investigates CIA as a current source of suffering in the horse and a growing threat as climate change raises ambient temperatures around the globe. We have already completed analysis of banked genome sequences, identifying novel polymorphisms in candidate genes. Most recently, we conducted a GWA study with 99 cases and 167 controls utilizing a threshold CIA score. GWAs revealed at least one locus of major effect (raw p< Bonf. Threshold of 1.18e-7). The genomic heritability for this trait is estimated at 84%. While these findings support the hypothesis that a hereditary or genetic component is contributing to this condition, more work is current ongoing to narrow down the causative loci and intervals. Although CIA is clearly a multifactorial condition and likely results from the action of several genes, our genomic approaches will provide information on genetic factors underlying etiological pathways controlling this condition.

P0365: Equine

SNP-Based Heritability and Genetic Architecture of Hock Osteochondrosis in Standardbred Horses

Annette M McCoy[1], Elaine M Norton[2], Ann M. Kemper[1], Samantha K Beeson[2], James R Mickelson[2] and Molly E McCue[2], (1)University of Illinois, Urbana, IL, (2)University of Minnesota, St. Paul, MN
Osteochondrosis (OC) is a common developmental orthopedic disease characterized by a failure of endochondral ossification. Certain joints, and locations within these joints, are predisposed to the development of OC, not only across horse breeds, but across species. There is strong

evidence for a heritable component to this disease, but estimates of heritability vary widely based on population and disease definition. Prior heritability estimates for OC in horses have been based on pedigree data or low-density genotyping data, and have ranged from 15-52%. The aim of this study was to use high-density SNP data to estimate heritability and genetic architecture in a large cohort of Standardbreds stringently phenotyped for hock OC.
Genome-wide genotyping was performed in 479 North American Standardbred horses (n = 148 cases; n = 331 controls). Genotyping data was imputed to ~2 million single nucleotide polymorphisms (SNPs) using an established pipeline. Imputed genotypes were imported into BayesR for estimation of genetic architecture. SNP-based heritability of OC in this population was explained by 2326 SNPs. The majority of these SNPs (86.6%) had small effects, as expected for a polygenic disease. Ten percent of SNPs had a moderate effect, while few SNPs (2.9%) had large effects. To determine heritability, a linkage-disequilibrium-weighted relatedness matrix was calculated in LDAK and utilized in restricted maximum likelihood (REML) analyses in both LDAK and GCTA software with sex as a covariate. Estimated heritability was 0.24 (SE 0.16, $p = 6.6 \times 10^{-4}$) in GCTA and 0.17 (SE 0.01, $p = 1.5 \times 10^{-3}$) in LDAK.

P0366: Equine
Genomic Variants Associated with Conformation and Craniofacial Features in Arabian Horses
Felipe Avila[1], James R Mickelson[2] and Molly E McCue[2], (1)University of Minnesota, Saint Paul, MN, (2)University of Minnesota, St. Paul, MN
The Arabian is a unique horse breed due to its morphological and cranial features. These horses are known for their chiseled head; also, some individuals have 5 lumbar vertebrae instead of 6 and 17 pairs of ribs instead of the usual 18. Because of that, Arabians have a shorter back and a more horizontal pelvis. The objectives of this study were to 1) identify candidate genes for cranial and body morphology within genomic signatures of selection, and 2) identify and annotate biologically relevant alleles within these genes to investigate their association with the Arabian's distinctive cranial morphology and body conformation. For that, 36 Arabians were genotyped at 2 million SNP markers across the genome, and regions harboring signatures of selection were identified. A total of 361 genes were retrieved within signatures of selection from 28 autosomes, prioritized using Endeavour-GW, and functional and enrichment analyses were performed using DAVID and FunRich3. Then, WGS data from 14 Arabians were used for variant calling using SamTools, and genomic variants were annotated using SnpEff. Currently, the top 50 candidate genes for the aforementioned traits of interest have been analyzed, and moderate and high impact variants were identified in several of these candidate genes. Next, we aim to validate these variants in an effort to characterize their role in determining the unique Arabian conformation and craniofacial features.

P0367: Other Animal Species
Creating a Multi-Organism Disease Model Resource at RGD
Mary Shimoyama, Medical College of Wisconsin, Milwaukee, WI
The Rat Genome Database (RGD) was initiated in 1999 to standardize, integrate and present genomic and phenotype data for the laboratory rat. Because of its use as a model for multiple human diseases and the needs of a disease focused community of researchers, RGD has included human and mouse data from its beginning including genes, QTL, ClinVar variants, and functional data such as disease, phenotype, pathway and Gene Ontology annotations. With the rise of precision medicine initiatives, the need for researchers to access and compare genomic and phenotype data for a variety of models ideal for particular disease studies has increased. To accommodate these needs, RGD adapted its data formats, technical infrastructure and data mining and presentation tools to accommodate data from organisms that serve as important models for specific diseases. Each organism has a genome browser and users can access multiple organism data in the InterViewer for protein-protein interactions and the Variant Visualizer. The Object List Generator & Analyzer (OLGA) and the Gene Annotator allow users to create data sets based on genome region, functional information and combinations of function and then analyze data for genome location, variant pathogenicity and functional commonalities. RGD currently includes data for chinchilla, dog, bonobo and 13 lined ground squirrel. RGD's flexible infrastructure will easily accommodate its planned continuous expansion to multiple other disease model organisms.

P0368: Other Animal Species
Functional Effects of a Retained Ancestral Polymorphism in Prestin
Yuanyuan Li, kunming institute of zoology, Kunming, China
Molecular basis for mammalian echolocation has been receiving much concerns. Recent findings on the parallel evolution of *prestin* sequences among echolocating bats and toothed whales suggest that adaptations for high-frequency hearing have occurred during the evolution of echolocation. Here, we report that although the species tree for echolocating bats emitting echolocation calls with frequency modulated (FM) sweeps is paraphyletic, *prestin* exhibits similar functional changes between FM bats. Site-directed mutagenesis shows that the amino acid 308S in FM bats is responsible for the similar functional changes of *prestin*. We strongly support that the occurrence of serine at position 308 is a case of hemiplasy, caused by incomplete lineage sorting of an ancestral polymorphism. Our study not only reveals sophisticated molecular basis of echolocation in bats, but also calls for caution in the inference of molecular convergence in species experiencing rapid radiation.

P0369: Other Animal Species
Integrative Biobanking to Identify and Interpret Functional Genomic Data from Rhesus Macaques on Cayo Santiago Island, Puerto Rico
Michael J. Montague[1], Noah Snyder-Mackler[2], Lauren J.N. Brent[3], J.H. Skene[4], Julie E. Horvath[5,6] and Michael L. Platt[1], (1)University of Pennsylvania, Philadelphia, PA, (2)University of Washington, Seattle, WA, (3)University of Exeter, Exeter, United Kingdom, (4)Duke University, Durham, NC, (5)North Carolina Central University, Durham, NC, (6)North Carolina Museum of Natural Sciences, Raleigh, NC
Evidence suggests that individual variation in social behavior arises from a combination of genetic predispositions and individual experience, yet the underlying biological mechanisms remain poorly understood. To address this gap, we have sought to understand the genetic contributions to social behavior in a large, free-ranging population of rhesus macaques (*Macaca mulatta*) with a known pedigree and detailed

behavioral phenotypes. We hypothesize that genetic variants underlying molecular differences in neural circuits are associated with behavioral variation in this socially complex species. Toward this end, over the past seven years, we have compiled genetic data and biological samples from approximately 1000 animals, while also amassing extensive behavioral and cognitive data through assays and focal observations. After conducting whole genome sequencing for 250 individuals, we performed additional tissue sampling to biobank brain tissue and twenty peripheral tissues and organs sampled from a subset of animals, as well as microbiota from skin, lung and digestive tract. Whole blood and flash frozen brain tissues were recently processed for gene expression (RNAseq) and epigenetic (ATACseq) analyses. The brain bank is unprecedented in the number, diversity of ages, and degree of background behavioral information, and this multi-faceted approach will generate valuable insights on the influences of social environment, including not only social stressors, but also social mechanisms that may be protective against stress. Here, we illustrate how this integrative dataset can be used to identify the functions and pathways underlying the relationship between social environment and the transcriptional and epigenomic signatures in different brain regions.

P0370: Other Animal Species
Low-Coverage Genome Sequencing of Black Rhinoceros to Inform Conservation Management
Sergio A Redondo, Stanford University, Stanford, CA
Anthropogenic pressures have tremendously reduced the black rhinoceros (*Diceros bicornis*) population and distribution. Currently just over 5,000 individuals survive in protected African reserves. While conservation efforts have been effective, the species remains critically endangered and populations must be actively protected from poaching. Management techniques, including dehorning and translocation, have been used to protect individuals and populations, however, the efficacy of these invasive techniques has rarely been studied and the reproductive impacts have never been assessed. Small population sizes and low genetic diversity also raise important management issues. Because of their low genetic diversity, traditional measures of genetic diversity and structure may provide biased or uncertain estimates. Important population genetic parameters (e.g. effective population size, inbreeding) are not known with certainty, and hidden genetic structure (due to unknown relatedness) may be present. We are currently addressing these issues using low-coverage genome sequencing of one of the largest collections of rhino tissue samples collected in Africa (>800 individuals). It includes nearly all the founding individuals that were introduced to Lowveld conservancies, as well as most of the progeny born in these rapidly growing populations. In the first phase, we are sequencing 254 low coverage genomes. This will help to 1) determine parentage for offspring and evaluate the effect of management actions (e.g. dehornings, translocations) on reproductive success and breeding patterns and 2) determine relevant genetic characteristics of each population. Our results will be directly applicable to the genetic and demographic management of free-ranging populations and for planning rhino re-introduction projects.

P0371: Other Animal Species
Chromosome Level Assembly of the Water Buffalo Genome
Lloyd Low, Davies Research Centre, Roseworthy, Australia
Water buffalo (*Bubalus bubalis*), also known as Asian buffalo, is a globally important species for agriculture. A good reference genome of this species is necessary for understanding its biology, managing genetic diversity and to apply new genome-based selection methods for genetic improvement. We have previously reported a *de novo* assembled genome based on 454 and Illumina short reads. Due to the limitations of using only short reads in the assembly, the publicly available genome consists of 366,983 scaffolds, with a scaffold N50 of 1,412,388 bp and an L50 of 581 scaffolds. Here we report an improved diploid assembly of the same individual (Olimpia) using PacBio Sequel and RSII reads (> 69-fold coverage), assembled with the FALCON-Unzip assembler. Despite using a highly inbred animal with a high homozygosity level, 58% of its genome could be phased with a haplotig N50 of 0.394 Mbp and the longest haplotig is 2.77 Mbp. Scaffolding using Chicago and HiC reads was done using the Dovetail HiRise software along with custom scripts to remove false contig breaks that were artificially introduced. Improvement in the latest assembly scaffold N50 is 117,187,264 bp which is 83-fold longer than the previous short read based assembly, and has an L50 that is made up of only 9 scaffolds. The current draft of the assembly is still evolving as we continue to correct potential mis-orientation within scaffolds and examine conservation of synteny with the cattle and goat genomes, before final gap filling, assembly polishing and annotation.

P0372: Other Animal Species
Association between Expression of the Genes Related to Subclinical Mastitis and Milk Components in Water Buffalo
Fernanda Tanamati, Daniel Jordan de Abreu Santos, Nedenia Bonvino Stafuzza, Daniele Fernanda Jovino Gimenez and Humberto Tonhati, São Paulo State University - UNESP, Jaboticabal, Brazil
Mastitis in water buffalo causes economic losses due to decreased milk production, making it less suitable for consumption and processing, along with having a negative effect on the animals' welfare. The aim of this study was to verify the relationship between the expression of the lactoferrin (*LTF*), tumor necrosis factor alpha (*TNF-α*), interleukin-1 beta (*IL-1β*), interleukin-8 (*IL-8*) and toll-like receptors 2 (*TLR-2*) and 4 (*TLR-4*) genes with the somatic cell score (SCS), fat, protein, lactose, total solids (TS) and solids-not-fat (SNF) components, in healthy and affected subclinical mastitis buffaloes. Twelve milk samples with and 12 milk samples without subclinical mastitis, were collected. To determine which animals presented subclinical mastitis we used the California Mastitis Test indirect method (weakly positive, positive, strongly positive, and negative) and the Somatic Cell Count direct method (without mastitis $\leq 161 \times 10^3$ cells/mL and mastitis $\geq 860 \times 10^3$ cells/mL). Linear regression analysis was performed to determine the relationship between ΔCt of the genes and the studied components. The highest coefficients of determination were observed between SCS and almost genes, except for the *LTF* gene. *TLR-2* is the gene which its expression most explains the variation of SCS. Linear regression coefficient was significant for SCS and *IL-1β, IL-8, TLR-2, TLR-4* and *TNF-α* genes. *TNF-α* gene obtained the highest linear regression coefficient. Our results show a greater relation of the studied genes with the SCS, and indicate that the *TLR-2* and *TNF-α* genes are highly associated with this component.
Acknowledgments: São Paulo Research Foundation (FAPESP) grant numbers #2014/19321-4 and #2016/10526-8.

P0373: Other Animal Species
Allele–Specific Expression of Macrophage Expressed Genes in Water Buffalo

Prasun Dutta[1], Rachel Young[1], Lucas Lefèvre[1], Stephen J. Bush[1], Eileen Wall[2] and David A. Hume[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Scotland's Rural College (SRUC), Edinburgh, United Kingdom

Tuberculosis or TB (caused by *Mycobacterium bovis*) and brucellosis (caused by *Brucella abortus*) are zoonotic diseases affecting livestock species, including water buffalo (*Bubalus bubalis*). This project aims to identify and validate polymorphisms in genes associated with susceptibility and resistance to TB and brucellosis in Indian domestic water buffalo. As both pathogens replicate inside macrophages, genes associated with differences in susceptibility are likely to be expressed specifically in these cells. Genes with enriched expression in macrophages were identified using an atlas of gene expression, created using RNA-sequencing data and spanning multiple tissues and cell types. DNA and RNA sequence data was used to identify protein-coding variation in macrophage-expressed genes, including potential null mutations. Variants that affect the expression level of macrophage-enriched genes were detected by their allelic imbalance; that is, unequal levels of transcription from two alleles. The extreme of allelic imbalance, monoallelic expression, is when one allele is not expressed at all. This could arise by imprinting or a loss of function. In these cases, we examine the corresponding genomic sequence to explore the likely mechanism. Overall, this work will explore the prevalence of allele-specific expression in buffalo macrophages, and show how allelic imbalance can help identify *cis*-acting regulatory variations in candidate macrophage-expressed genes.

P0374: Other Animal Species

A *de novo* Hybrid Assembly of a Dromedary Camel

Heather M. Holl[1], Donald Miller[2], Salma Abdalla[3], Benjamin M. Shykind[3], Joel A. Malek[3], Yasmin A. Mohamoud[4], Ayeda Ahmed[5,6], Kamaal Pasha[7], Adel Khalili[7], Douglas Antczak[2] and Samantha A. Brooks[1], (1)University of Florida, Gainesville, FL, (2)Cornell University, Ithaca, NY, (3)Weill Cornell Medical College, Doha, Qatar, (4)WCMC-Qatar, Doha, Qatar, (5)Weill Cornell Medical College in Qatar, Doha, Qatar, (6)Weill Cornell Medical College in Qatar, Qatar, (7)Tharb Camel Hospital, Doha, Qatar

The dromedary camel is of great economic and cultural importance in many countries, often selected for meat and milk production, draught, riding, and racing traits. The global population is estimated to be over fifteen million, many of which are in developing nations. Camels possess an array of unique physiological adaptations that have enabled their survival in an arid desert environment. The two currently available dromedary reference genomes were assembled using only Illumina short reads, containing roughly thirty thousand scaffolds with a N50 of single digit megabases. Our goal was to produce an assembly incorporating long read technologies to improve contiguity, which will facilitate genomic selection and a comparative genomic study of desert adaptation in mammals. We selected a male dromedary camel from the US to serve as our genome reference, establishing a fibroblast cell line as a continuous source of high molecular weight DNA. Hybrid assembly incorporated 74x of paired end Illumina reads and 15x PacBio long reads. Scaffolding was accomplished by incorporating 30x coverage of 10X Genomics Chromium sequencing. Scaffolds were organized into putative chromosomes with comparative genomics using the RACA method. Resulting chromosomes were evaluated and assigned using alpaca genetic maps and dromedary radiation hybrid maps. A *de novo* transcriptome assembly was available to assist with gene annotation. The resulting assembly is a valuable resource for future genomic studies in the dromedary camel.

This study was made possible in part by NPRP Grant 6-1303-4-023 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors.

P0375: Other Animal Species

Generating Y-Chromosomal Shotgun Assemblies for Old World Camelids to Study Their Male Genealogies

Sabine Felkel[1], Barbara Wallner[1], Hussain Bahbahani[2], Faisal Almathen[3] and **Pamela Burger**[4], (1)Institute of Animal Breeding and Genetics, Vetmeduni Vienna, Vienna, Austria, (2)National Unit for Environmental Research and Services (NUERS), Faculty of Biological Sciences, Kuwait University, Kuwait, Kuwait, (3)King Faisal University, Al-Hasa, Saudi Arabia, (4)Vetmeduni Vienna, Research Institute of Wildlife Ecology, Vienna, Austria

Polymorphic on the markers on the male specific part of the Y-chromosome (MSY) provide useful information for tracking male genealogies. While maternal lineages are well studied in Old World camelids using mtDNA, the lack of Y-chromosomal references hampers the analysis of male driven demographies. Recently it has been shown in horses, that a shotgun assembly generated from short read next generation sequencing (NGS) data revealed sufficient resolution to trace individual male lines in this species. In a similar approach we generated MSY shotgun assemblies for *Camelus dromedarius* and *Camelus bactrianus* by *de novo* assembling NGS data enriched for Y-chromosomal reads followed by a remapping-filtering approach. The resulting assemblies are used as a reference for variant calling using short-read data from multiple individuals and to generate first Y-chromosomal phylogenies for these species.

P0376: Other Animal Species

Mitochondrial Diversity and Phylogeographic Structure of Domestic Bactrian Camel

Rendalai Si[1], Haiqing Liu[1] and Rimutu Ji[1,2], (1)Inner Mongolia Institute of Camel Research, Alxa, Inner Mongolia, China, (2)Inner Mongolia Agricultural University, Hohhot, Inner Mongolia, China

Aims: China is a country with the richest Bactrian camel resources all over the world, and there are very few parts of the wild Bactrian camels. Bactrian camel once played an important role in culture exchange, economic development and desert animal husbandry in the history. However, but so far there has been no extensive study on genetic diversity, population demographic history, and origin of Chinese breeds.
Methods: In this study, two kind of gene sequences including 809-bp fragment and 1140-bp fragments of the *cytochrome b* gene (*Cytb*) of mitochondrial DNA (mtDNA) from 138 individuals, representing 13 domestic Bactrian camel populations, including Chinese, Mongolian, Russian and Kazakhstan were amplified by polymerase chain reaction and sequenced directly. Sequences of the partial mtDNA were aligned using MEGA; the parameters of the mitochondrial polymorphisms were computed using DNASP, including nucleotide diversity (p), the number of haplotypes, haplotype diversity (Hd) and its standard error (SE); A median-joining network was drawn using the program NETWORK to investigate possible relationships among the haplotypes; phylogenetic trees were constructed using the neighbourjoining

method. **Results:** Based on the 809-bp fragment of mtDNA, we found 17 different haplotypes, the number of haplotypes detected in each domestic populations varied from three to eight, and the haplotype diversity values ranged from 0.8916±0.2376 to 0.5782±0.1826. The nucleotide diversity (p = 0.0081) and haplotype diversity (Hd = 0.826±0.022) of all the Bactrian camels in this study were calculated. Among the 20 different haplotypes identified from the 138 *Cytb* sequences, we identified 21 variable sites (20 transitions and 1 transversion). We found that the number of haplotypes from different populations varied from 3 to 8, and haplotype diversity values from 0.5411 in the Sunit to 0.9131 in the Tarim Bactrian camel breed. The Tarim Bactrian camel displayed the highest nucleotide diversity value (0.00151), whereas the Sunit breed had the lowest value (0.00021). **Conclusion:** The analysis of domestic Bactrian camels from different geographical locations found there was no significant genetic divergence in China, Russia, Mongolia and Kazakhstan. This suggests a strong gene flow due to wide movement of domestic Bactrian camels.

Key Words: Cytochrome b gene (Cytb); Domestic Bactrian camel; Bactrian camel breed; Haplotype;
The work was supported by International S & T Cooperation Program of China (2015DFR30680 and ky201401002); and Chinese National Natural Science Foundation project (31360397).

P0377: Other Animal Species

Identification of Genetic Variants Affecting Color and Phenotypic Variation between Dromedary Populations using Whole Genome Sequencing (*Camelus dromedarius*)

Fahad Alshanbari, Texas A&M University, College Station, TX

Coat color serves many purposes in the wild, but is also a target of selection by humans. The genetic mechanisms underlying coat color vary by species, but are typically regulated by few major genes and further modified in pattern and intensity by other genes, which ultimately regulate levels and distribution of eumelanin and pheomelanin. The melanocortin 1 receptor (MC1R) and agouti signaling protein (ASIP) are associated with coat color variation, but nothing has yet been documented in the dromedary. Recently, we showed that a nonsense mutation in *ASIP* is associated with black, and a missense mutation in *MC1R* with the white color in the dromedary. TaqMan assay analysis in large cohorts suggests additional modifier genes to produce a range of different shades of black/brown/red/beige/white. We pooled DNA by color phenotype from 17 black, 27 white and 31 red Saudi Arabia dromedaries and sequenced by Illumina NextSeq500. Using a sliding window zF_{ST} outlier approach combined with intra-population nucleotide diversity, we observed high F_{ST} ($Z>8$) and low nucleotide diversity surrounding *ASIP* in black compared to white and red dromedaries, confirming our prior findings. However, we did not observe this pattern surrounding *MC1R*, perhaps due to lack of strong selection or sequence complexity. Additionally, 150 putative selective sweep regions were identified, indicating that these moderately isolated populations retain genomic differentiation despite some gene flow. The artificially imposed demography leads to differences related to production traits such as milk and meat production, representing genetic resources useful for producers to assist in breeding programs.

P0378: Other Animal Species

Resequencing of Reindeer Genomes: Insights into Genetic Diversity, Adaptation and Selection Signatures

Melak Weldenegodguad[1,2], Knut Røed[3], Kisun Pokharel[1], Mervi Honkatukia[1], Tiina Reilas[1], Jaana Peippo[1] and Juha Kantanen[1], (1)Natural Resources Institute Finland, Jokioinen, Finland, (2)University of Eastern Finland, Kuopio, Finland, (3)Norwgian University of Life Sciences, Oslo, Norway

Reindeer (*Rangifer tarandus*) have adapted to the northern and subarctic regions (between 50 – 81 degrees of latitude) in the Eurasian and North American regions where this semidomesticated species has pivotal economic, societal and cultural values for indigenous societies. We performed a whole genome resequencing of 24 semi-domesticated and wild reindeer samples originating from Russia, Norway, Finland and USA to investigate genetic diversity, interpopulation relationships and selection signatures. We reached 10-fold coverage after mapping the sequencing reads on the draft reindeer genome (Abstract ID #28814). More than 19 million SNPs were identified in the samples. Reindeer displayed heterozygosity rate of 1.3X10e-3, which is 2.3 and 1.5 times higher than what is found in taurine cattle (0.59X10e-3) and yak (0.89X10e-3). The higher level of genetic variation in reindeer may be associated with gene flow from wild reindeer populations, less intensive artificial selection practices and shorter domestication history. In the analysis of phylogenetic relationships, the sequenced animals grouped into two main clusters: northern European and northern Russian/northern American. The pattern may reflect differences in the colonization history of the two geographic clusters after the Last Glacial Maximum. The northern Russian/northern American cluster may have descended from populations spread from the Beringia refugium meanwhile the northern European reindeer originated from the European refugium. The finding suggests that there have been at least two domestication events in the history of reindeer.

P0379: Other Animal Species

Sequencing and *De Novo* Assembly of a Reindeer (*Rangifer tarandus*) Genome

Kisun Pokharel[1], Yao Ming[2], Melak Weldenegodguad[1,3], Mervi Honkatukia[1], Tiina Reilas[1], Jaana Peippo[1] and Juha Kantanen[1], (1)Natural Resources Institute Finland, Jokioinen, Finland, (2)BGI-Shenzhen, Shenzhen, China, (3)University of Eastern Finland, Kuopio, Finland

Reindeer (*Rangifer tarandus*) is a culturally and economically important semi-domesticated species in the northern and subarctic regions in Eurasia and North America. A draft quality of reference genome will be valuable to study the domestication history and get insight into the genetic basis of adaptation. Moreover, having a reference genome will facilitate selection approaches and establishing genetic conservation programmes for these vulnerable animals taking into consideration the effect of climate change.

Here we report a draft reference assembly and annotation of a Finnish male reindeer genome. Using Illumina's HiSeq 4000 and HiSeq 2000 platforms, we generated 512.9 Gb of raw data from a series of paired-end and mate-pair libraries. A total of 300.5 Gb of clean data was assembled using SOAPdenovo resulting into 256,454 scaffolds (N50 = 502 Kb) with cumulative scaffold length of 2.66 Gb and spanning 90% of the estimated (2.9 Gb) genome size. Using homology based approach, the reindeer genome was predicted to harbour 28,212 protein coding genes, 97.97% of which have been functionally annotated. BUSCO evaluation demonstrated that the reindeer genome gene set assembly was 95% complete. The mitochondrial genome of reindeer (16,451 bp) comprises 14 protein-coding genes, 22 tRNAs and two 2 rRNAs. The

genome itself and orthologous comparison with other mammalian species provided important insights into the evolution and demographic history of the reindeer.

P0380: Other Animal Species

Chromosomal Evolution of Muntjac Deer Revisited By High Quality Reference Genomes

Yuan Yin, Northwestern Polytechnical University, Xi'an, China

The *Muntiacus* genus in the Cervidae is a good model for studying chromosomal evolution because it has extremely diverse chromosome number variation caused by recurrent tandem fusion events. Here, we report high-quality genomes of a male black muntjac (*Muntiacus crinifrons*, 2=8/9) assembled to the chromosome level, and a female Chinese muntjac (*Muntiacus reevesi*, 2n=46) and an Indian muntjac (*Muntiacus muntjak*, 2n=6/7) assembled to the draft level respectively. The genome sequences of the black muntjac assembled using PacBio and HiC technologies read through several chromosomal fusion sites and even a few centromere regions rich in satellite repeats. We observed a number of DNA repair genes had evolved rapidly only in the *Muntiacus*. All these data will help us to have a better understanding about the underlying mechanism of tandem fusion in this genus. There is also a neo-sex chromosomes system in the black muntjac which is the youngest neo-sex chromosomes system identified in mammals. We found slight degeneration on the neo-Y chromosomal region and tendency of dosage compensation. As the first chromosomal level assembly of genome in the Cervidae family, the genome of the black muntjac, together with other two congenic species' reference genomes, the results obtained in this study provide excellent sources to study the unusual chromosomal evolution of *Muntiacus* as well as good model for investigating sex chromosomes evolution of mammals.

P0381: Other Animal Species

Comparative Genomics and Genome Evolution in Birds of Paradise

Stefan Prost, Program for Conservation Genomics, Stanford University, Stanford, CA

The diverse array of phenotypes and lekking behaviors in Birds-of-paradise have long excited scientists and laymen alike. In order to understand the genomics underlying the evolution of this remarkable groups of birds, we sequenced genomes representing species from all five clades within Birds-of-paradise. GO enrichment of positively selected genes on the branch to the Birds-of-paradise show an enrichment for collagen and eye developmental genes. We further found genes important for coloration and beak development to be in the set of positively selected genes. Startle response (response to predators) was enriched among gene families that evolve significantly faster in the core Birds-of-paradise compared to other birds in our study. Furthermore, we found a burst in repeat activity in all our genomes dating to around the time of the diversification of the Bird-of-paradise group, which could potentially play a role in the evolution of this fascinating group of birds.

P0382: Other Animal Species

New World Quail Population Genomics Reveals Structure, Divergence, and Evidence for Selection

Miranda Loraine Wilson[1], David L Oldeschulte[1], Yvette A Halley[2], Eric Bhattarai[1], Wesley Brashear[1], Josh Hill[1], Richard P Metz[3], Charles Johnson[1], Dale Rollins[4], Markus Peterson[5], Derek M. Bickhart[6], Jared E. Decker[7], John F. Sewell[1] and Christopher M. Seabury[1], (1)Texas A&M University, College Station, TX, (2)West Virginia University, Morgantown, WV, (3)AgriLife Genomics and Bioinformatics Services, College Station, TX, (4)Texas A&M University, San Angelo, TX, (5)The University of Texas at El Paso, El Paso, TX, (6)Dairy Forage Research Center, USDA-ARS, Madison, WI, (7)Division of Animal Sciences, University of Missouri, Columbia, MO

North American quail species have been a focus of wildlife researchers for decades, primarily due to more recent broad-scale population declines. The scaled quail (*Callipepla squamata*) is a desert niche species, whose range extends across the southwestern United States and into Mexico, and also overlaps in Texas with the range of the more widely distributed and phenotypically diverse northern bobwhite (*Colinus virginianus*). A first-generation draft genome assembly for the scaled quail, and a second-generation draft genome assembly for the northern bobwhite, have recently been published. We annotated these genomes using a combination of cDNA and protein evidence from the northern bobwhite (*Colinus virginianus*), chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and turkey (*Meleagris gallopavo*). These assemblies are currently being used for northern bobwhite and scaled quail population genomics, with a primary focus on population structure, divergence, and the detection of genomic regions that may be subject to intense selective pressures. To date, we have identified and are examining thousands of genomic regions that do not follow a neutral model, based on the frequency distribution(s) of predicted polymorphisms for 57 scaled quail and 126 northern bobwhites that were sampled from diverse ecoregions across the southern United States of America. We hypothesize that the underlying biological function(s) of genes and/or regulatory features present among loci putatively subject to intense selective pressures may potentially reveal new insights regarding some factors contributing to modern quail declines.

P0383: Other Animal Species

Genome Sequence of the Pond Wolf Spider, *Pardosa pseudoannulata*

Zewen Liu, Nanjing Agricultural University, Nanjing, China

The powerful venom and excellent silk make spiders to be superb predators against a range of insect pests. Here, we present the assembled and annotated genome of the pond wolf spider, *Pardosa pseudoannulata*, an important predator enemy through most Asian countries and areas. The genome of *P. pseudoannulata* is larger, close to 4.5 G, than that of *Stegodyphus mimosarum* and *Nephila clavipes*, and nearly half of the DNA sequences are repetitive, which made the genome assembly a big challenge. However, by employing different sequencing strategies together with deep sequencing, the genome is assembled with the contig N50 of 22 Kb and scaffold N50 of 457Kb, which provides solid foundation to the genome annotation. More than twenty thousand genes are annotated covering 90% of total genes. The analysis and diversity of genes involved in the germination of silk and venom and other interesting families will provide useful information for basic studies on the arthropod evolution and for application researches as a predator enemy. For example, on the cholinergic signaling system of *P. pseudoannulata*, there are 17 nAChR subunits (12 α subunits and 5 β subunits), with the most number of β subunits among arthropods. There are 5 AChBPs which are not found in the insects and acarina. At least 7 acetylcholinesterases were found in *P. pseudoannulata*, which is the most in number by far.

We are now seeking cooperation opportunities with senior researchers on molecular and evolutional studies from both whole genome and protein family aspects.

P0384: Other Animal Species

***De novo* Assembly of the Tube-Snout Genome and Comparative Genomics Study**

Qiushi Li[1,2] and Sam Yeaman[1], (1)University of Calgary, Calgary, AB, Canada, (2)China Academy of Chinese Medical Sciences, Beijing, China

Here we report the 487.86 Mb genome assembly of the tube-snout (*Aulorhynchus flavidus*), a marine fish naturally found along the shallow waters off the Pacific coast of North America. The tube-snout with an elongated body shape is closely related to the sticklebacks, which have an extensive history of genetic studies. The haploid contigs of the tube-snout genome with N50 > 1.8 Mb were assembled from 100x Pacbio long-reads by the diploid aware genome assembler "Falcon unzip". Dovetail HiRise pipeline was used to augment the draft to a chromosome-level assembly with scaffold N50 > 18.4 Mb spanning ~97% of the estimated genome size. 16.9 Kb mitochondrial genome sequence was circularized without gaps. In addition, we characterized the genomic sequences of the tube-snout. A comprehensive comparison between the tube-snout genome and the newly released high-quality three-spined stickleback (*Gasterosteus aculeatus*) genome was performed. The interpretation of the tube-snout genome provides the foundation to identify how commonly genomic rearrangements have contributed to the evolution of clustered genome architectures among vertebrates.

P0385: Other Animal Species

Draft Genome of the Sea Cucumber *Apostichopus japonicus* and Genetic Polymorphism Among Color Variants

Jihoon Jo, Jooseong Oh, Hyun-Hee Hong, Sung-Gwon Lee, Seongmin Cheon and Chungoo Park, Chonnam National University, Gwangju, South Korea

The Japanese sea cucumber (*Apostichopus japonicus* Selenka 1867) is an economically important species as a source of seafood and ingredient in traditional medicine. It is mainly found off the coasts of northeast Asia. Recently, substantial exploitation and widespread biotic diseases in *A. japonicus* have generated increasing conservation concern. However, the genomic knowledge base and resources available for researchers to use in managing this natural resource and to establish genetically based breeding systems for sea cucumber aquaculture are still in a nascent stage. A total of 312 Gb of raw sequences were generated using the Illumina HiSeq 2000 platform and assembled to a final size of 0.66 Gb, which is about 80.5% of the estimated genome size (0.82 Gb). We observed nucleotide-level heterozygosity within the assembled genome to be 0.986%. The resulting draft genome assembly comprising 132 607 scaffolds with an N50 value of 10.5 kb contains a total of 21 771 predicted protein-coding genes. We identified 6.6–14.5 million heterozygous single nucleotide polymorphisms in the assembled genome of the three natural color variants (green, red, and black), resulting in an estimated nucleotide diversity of 0.00146. We report the first draft genome of *A. japonicus* and provide a general overview of the genetic variation in the three major color variants of *A. japonicus*. These data will help provide a comprehensive view of the genetic, physiological, and evolutionary relationships among color variants in A. japonicus, and will be invaluable resources for sea cucumber genomic research.

P0386: Other Animal Species

New Insights into Global Biogeography, Population Structure and Natural Selection from the Genome of the Epipelagic Copepod Oithona.

Mohammed-Amin Madoui[1], Julie Poulain[1], Kevin Sugier[2], Benjamin Noel[2], Karine Labadie[1], Astrid Cornils[3], Leocadio Blanco-Bercial[4], Jean-louis Jamet[5] and Patrick Wincker[1], (1)CEA - Genoscope, Evry, France, (2)CEA-Genoscope, Evry, France, (3)Alfred Wegener Institute, Bremerhaven, Germany, (4)Bermuda Institute of Ocean Sciences, Bermuda, Bermuda, (5)Université de Toulon, Toulon, France

In the epipelagic ocean, the genus *Oithona* is considered as one of the most abundant and widespread copepods and plays an important role in the trophic food web. Despite its ecological importance, little is known about *Oithona* and cyclopoid copepods genomics. Therefore, we sequenced, assembled and annotated the genome of Oithona nana. The comparative genomic analysis integrating available copepod genomes highlighted the expansions of genes related to stress response, cell differentiation and development, including genes coding Lin12-Notch-repeat (LNR) domain proteins. The *Oithona* biogeography based on 28S sequences and metagenomic reads from the *Tara* Oceans expedition showed the presence of *O. nana* mostly in the Mediterranean Sea (MS) and confirmed the amphitropical distribution of *Oithona similis*. The population genomics analyses of *O. nana* in the Northern MS, integrating the *Tara* Oceans metagenomic data and the *O. nana* genome, led to the identification of genetic structure between populations from the MS basins. Furthermore, 20 loci were found to be under positive selection including four missense and eight synonymous variants, harbouring soft or hard selective sweep patterns. One of the missense variants was localized in the LNR domain of the coding region of a male-specific gene. The variation in the B-allele frequency with respect to the MS circulation pattern showed the presence of genomic clines between *O. nana* and another undefined *Oithona* species possibly imported through Atlantic waters. This study provides new approaches and results in zooplankton population genomics through the integration of metagenomic and oceanographic data.

P0387: Other Animal Species

The Developmental Transcriptome Atlas of the Spoon Worm *Urechis unicinctus* (Echiurida: Annelida)

Sung-Gwon Lee, Jooseong Oh and Chungoo Park, Chonnam National University, Gwangju, South Korea

Background: Echiurida is one of the most intriguing major subgroups of the phylum Annelida, because unlike other annelid members, most echiurids lack metameric body segmentation as adults. For this reason, transcriptome analyses from various developmental stages of Echiurida species can be of substantial value for understanding how gene repertoires are involved in early stages of development,ontogenic morphogenesis and the formation of an unsegmented body plan.

Finding: A total of 914 million raw RNA-Seq reads were produced from 14 developmental stages of *Urechis unicinctus*, and were *de novo* assembled into contigs spanning 63,928,225 bp with an average length of 1,481 bp. The resulting comprehensive transcriptome database of the

early developmental stages of *U. unicinctus* consists of 25,986 representative protein-coding functional transcripts. Approximately 66 % of unigenes were assigned to superphylum-level taxa, including Lophotrochozoa (40%). The completeness of the transcriptome assembly was assessed using BUSCO, and 71.3 % of the metazoan single-copy orthologs were presented in our transcriptome database. We observed three distinct patterns of global transcriptome profiles from 14 developmental stages, and identified a total of 12,910 genes that showed dynamic regulation patterns during the differentiation and maturation of *U. unicinctus* cells.
Conclusions: We present the first large-scale developmental transcriptome dataset of *U. unicinctus* and provide a general overview of the dynamics of global gene expression changes during its early developmental stages These data are a first step toward understanding the complex developmental gene regulatory networks in *U. unicinctus,* and will furnish a valuable resource for analyzing the functions of gene repertoires in various developmental phases.

P0388: Other Animal Species
***De novo* Assembly of the *Schizocardium californicum* Genome using DNA Isolated from Sperm**
Gregory T Concepcion, PacBio, Menlo Park, CA; Pacific Biosciences, Menlo Park, CA
Efforts to sequence and assemble the genomes of highly heterozygous non-model organisms have proven to be difficult with traditional short read approaches. Long repetitive DNA structures as well as extensive structural variation between haplotypes in polyploid species coupled with a large genome size are limiting factors to achieving highly contiguous genome assemblies. Other factors paramount to generating a high quality reference are the quality and size distribution of the starting genomic DNA (gDNA) which is often difficult to obtain for non-model organisms due to co-purification of metabolites which can limit sequencing and assembly success. Here we demonstrate the utility of long DNA reads to generate a high quality *de novo* reference sequence from a sperm sample isolated from a highly polymorphic species. High quality gDNA isolated from hemichordate sperm circumvented many common DNA isolation pitfalls. Extracted gDNA was used to generate large insert (>30 kb) libraries for subsequent SMRT Sequencing. Using Pacific Biosciences' Sequel System, the DNA was sequenced and a genome was assembled of approximately 1.6Gb with a contig N50 of ~739Kb. After 2 rounds of polishing with the Arrow consensus calling algorithm, 949 out of 978 (97%) BUSCO orthologs were detected, with 693 (70.9%) detected in duplicate, indicating assembly and resolution of different haplotypes in the primary assembled contigs. Animals in the phylum *Hemichordata* have provided key understanding of the origins and development of the vertebrate notochord. Here we present the highly contiguous *de novo* assembly and preliminary annotation of an indirect developing hemichordate genome, *Schizocardium californicum*.

P0389: Other Animal Species
The Genetic Architecture of Acute Heat Stress and Acute Oxidative Stress Response in the Nematode *C. remanei*
Christine H Oconnor[1], Kristin L. Sikkink[2], Janna L. Fierst[3], John Willis[1] and Patrick C. Phillip[1], (1)University of Oregon, Eugene, OR, (2)University of Minnesota, Minneapolis, MN, (3)University of Alabama, Tuscaloosa, AL
Abiotic stress response is a functionally important trait whose genetic basis is conserved throughout animals. Studies in model organisms, in particular the nematode *Caenorhabditis elegans*, have found that mutations in a small handful of genes affect an organism's ability to respond to a wide variety of abiotic stresses, including heat stress, oxidative stress and UV stress, among others. In this study, we used experimental evolution to dissect the genetic architecture of heat stress and oxidative stress resistance in populations of the genetically diverse nematode *Caenorhabditis remanei* evolved for either heat stress or oxidative stress resistance, and to test for evidence of pleiotropy in the response to selection. We used pooled population whole genome sequencing and transcriptome data from the ancestor and all evolved populations in our analyses. We find hundreds of genomic sites that show a response to selection for heat or oxidative stress resistance but no evidence for pleiotropy, in contrast what we expect from gene knockout studies in *C. elegans*. This lack of correlated response to selection is reflected in the pattern of changes in gene expression as well. Our results indicate that over an evolutionary timescale, many genes of small effect, as opposed to a few genes of larger effect, are selected for. We demonstrate that experimental evolution in a genetically diverse model organism provides a powerful framework for dissecting the genetic architecture of complex traits, deepening our understanding of such traits and providing guides for how they may evolve in other organisms.

P0390: Other Species
The Transition of Non-Human Genetic Variation Data from dbSNP to EVA
Gary Saunders, EMBL-EBI, Cambridge, United Kingdom and Tao Tao, NIH/NLM/NCBI, Bethesda, MD
The long-term commitment to provide stable identifiers for genetic variation data is of key importance so that newly discovered variants and alleles can be referenced in publications, cross-linked between databases and integrated with successive reference genome builds. The European Variation Archive (EVA, https://www.ebi.ac.uk/eva/) and dbSNP are the primary open repositories for archiving, accessioning, and distributing genomic variation data including single nucleotide variants as well as short insertion and deletions (indels). Advances in genome sequencing technologies have resulted in unprecedented growth of genetic variation data, prompting a review of roles and responsibilities.
In a new agreement between dbSNP and EVA, all non-human genetic variation data submissions will be solely handled by EVA, and dbSNP will support only human variation data. Accordingly, the EVA is now responsible for the generation and maintenance of reliable accessions for non-human genetic variation data. The NCBI's dbSNP is responsible for assigning stable identifiers to human genetic variation data only.
The EVA follows the same accessioning principles as employed by dbSNP: non-human variants submitted to the EVA are issued 'Submitted SNP' (ss) accessions, and these are periodically clustered to form 'Reference SNP' (rs) accessions. The EVA is fully committed to the continuation of all existing dbSNP ss and rs accessions.
This joint poster presents the current status of this transition, future plans, working timelines and the subtle differences in the submission and access to genetic variation data at EVA compared with dbSNP.

P0391: Other Species
A Whole Genome Association Study in Bailinggu (*Pleurotus tuoliensis*)

Yongping Fu, Engineering Research Center of Chinese Ministry of Education for Edible and Medicinal Fungi/Jilin Agricultural University, Changchun, China
Bailinggu (*Pleurotus tuoliensis*) is a major, commercially cultivated mushroom and widely used for nutritional, medicinal, and industrial applications. Here we report a high-quality genome of domesticated *P. tuoliensis* (Bailinggu) obtained by PacBio SMRT sequencing. The assembled genome size was 45 M, and 95% of the genome assembly was contained in 24 contigs (N50 = 2.8 M). A phylogenetic tree was constructed including *P. tuoliensis* and ten other fungal species. The estimated divergence times between *P. tuoliensis* and *P. eryngii*, *P. tuoliensis* and *P. ostreatus* are 6.6 and 20.1 million years ago (Mya), respectively. We then investigated genomic variation between domesticated and wild Bailinggu based on the genome-wide resequencing of 54 strains. Evidence for the molecular footprints of artificial selection in Bailinggu was found, and it revealed signals of selection in 170 genes of domesticated strains, several of which relate to cellulose catabolic process, nitrogen compound metabolic process and environmental adaptation. In addition, the integrative genomic and transcriptomic analyses of domesticated Bailinggu were used to identify genes crucial to fruiting body formation and response to cold stimulation. Our results provide insights into the genetic basis of Bailinggu diversity and domestication. These findings will improve future research on Bailinggu breeding and other closely related species.

P0392: Other Species

Sequencing and Assembling the Nuclear Genome of the Antarctic Psychrophilic Green Alga *Chlamydomonas* sp. UWO241: Unravelling the Evolution of Cold Adaption

Xi Zhang, Western University, London, ON, Canada
The Antarctic harbours a variety of algae that can withstand extreme cold but falter at warmer temperatures (psychrophiles), including the unicellular green alga *Chlamydomonas* sp. UWO241. Little is known, however, about the origins and evolution of psychrophilic algae, and their nuclear genomes remain largely unexplored. For my PhD, I propose to sequence and assemble the entire nuclear DNA of UWO241 using a next-generation sequencing (NGS) approach, and then employ these data to better understand the evolution of photopsychrophily. DNA sequencing technologies have undergone tremendous advancements in recent years, but assembling, annotating, and analyzing a nuclear genome is still a huge undertaking, especially for small laboratory groups, partly because many eukaryotic genomes are repeat rich and contain thousands of genes and introns. To characterize the UWO241 genome I will, firstly, develop an assembly pipeline for processing high-throughput DNA sequencing reads into genomic contigs. These contigs, alongside RNA-sequencing data, will then be fed into an annotation pipeline, which I will design based on the most up-to-date eukaryotic bioinformatics gene-profiling software. Computational analyses will be carried out on an in-house computer, which I have constructed for the Smith Lab, as well as on Western's supercomputing network SHARCNET. Lastly, I will perform a wide range comparative genomic analyses of the UWO241 genome with those of other model green algae, including *Chlamydomonas reinhardtii*. Preliminary data already suggest that the UWO241 genome is exceptional in many ways—including its size (>230 Mb) and gene copy number—and at least some of these features appear to have a fundamental role in surviving in the extreme cold.

P0393: General Comparative

The BIG Data Center's Database Resources for Plant & Animal Genome Studies

Zhang Zhang, BIG Data Center, Beijing Institute of Genomics, Beijing, China
Zhang Zhang, BIG Data Center, Beijing Institute of Genomics
The BIG Data Center at Beijing Institute of Genomics (BIG) of the Chinese Academy of Sciences provides freely open access to a suite of database resources in support of worldwide research activities in both academia and industry. With the vast amounts of omics data generated at ever-greater scales and rates, the BIG Data Center is continually expanding, updating and enriching its core database resources through big-data integration and value-added curation, including BioCode (a repository archiving bioinformatics tool codes), BioProject (a biological project library), BioSample (a biological sample library), Genome Sequence Archive (GSA, a data repository for archiving raw sequence reads), Genome Warehouse (GWH, a centralized resource housing genome-scale data), Genome Variation Map (GVM, a public repository of genome variations), Gene Expression Nebulas (GEN, a database of gene expression profiles based on RNA-Seq data), Methylation Bank (MethBank, an integrated databank of DNA methylomes), and Science Wikis (a series of biological knowledge wikis for community annotations). In addition, three featured web services are provided, viz., BIG Search (search as a service; a scalable inter-domain text search engine), BIG SSO (single sign-on as a service; a user access control system to gain access to multiple independent systems with a single ID and password) and Gsub (submission as a service; a unified submission service for all relevant resources). All of these resources are publicly accessible through the home page of the BIG Data Center at http://bigd.big.ac.cn.

P0394: General Comparative

Completing the Plant & Fungal Trees of Life: Overview and Bioinformatics Perspectives

Jan T. Kim, Royal Botanic Gardens, Kew, Richmond, United Kingdom
Evolutionary trees are powerful tools for prediction, species discovery, and conservation. While the backbones of the plant and fungal trees of life are relatively well understood, progress towards completing them is limited by lack of sequence data for numerous genera and the vast majority of species.

Synthesis of phylogenies from multiple loci across whole genomes enables reconstruction of evolutionary trees with unprecedented resolution. Targeted sequence capture provides a scalable route to obtain sequences from multiple loci, and is applicable to samples with heavily degraded DNA, such as that from herbarium and fungarium specimens.

The Plant and Fungal Trees of Life (PAFTOL) project at the Royal Botanic Gardens, Kew, applies targeted capture to produce multilocus phylogenies in which all genera in the plant and fungal kingdoms are represented. It targets 353 nuclear genes, and also aims to obtain sequences from the plastid genome. Its objective of maximising the number of sequenced genera implies limiting sequencing depth for each genus. This results in specific bioinformatics requirements including high accuracy in associating reads to targeted loci, and locus assembly working at low depth and in the presence of coverage gaps due to introns, while scalability to large numbers of reads is secondary.

PAFTOL's comprehensive investigation of phylogenetic relationships will provide a rich resource enabling the discovery and study of evolutionary patterns in the plant and fungal kingdoms, and a unifying framework for comparative research. The project is an essential step towards the compilation of genomic data for all known plant and fungal species.

P0395: General Comparative

Plant Genome Database Release 2.5: A Standardized Plant Genome Repository for 233 Species

Jongsun Park, Hong Xi and Yongsung Kim, InfoBoss Co., Ltd., Seoul, South Korea

Currently, more than 200 plant species have been sequenced and/or published; however, there is no central repository for plant genome sequences. NCBI genome database, as a general sequence repository, does not contain all published plant genomes with gene models (e.g. *Utricularia gibba*). Ensembl (Release 37) and Phytozome (v 12.1) are another plant genome repositories containing 45 and 82 genomes, respectively, which is much less than currently available plant genomes. In addition, a lot of re-sequencing projects including *Arabidopsis thaliana* (>1,135 genomes) and *Oryza sativa* genomes (>3,000 genomes) do not provide assembled genome sequences for understanding intraspecies divergences. To overcome these problems, we developed a standardized plant genome database (http://www.plantgenome.info/) for collecting all available plant genomes with gene annotation pipeline with InterProCan, identification and comparison of simple sequence repeats (SSRs). Moreover, several genome-wide analyses can be conducted on the web site with the aid of GlobalScrap®. The Plant Genome Database release 2.5 contains 1,446 plant genomes (233 species) and four red algal genomes as an outgroup have been collected from diverse sources including NCBI, Phytozome, Ensembl, and independent plant databases as well as have been analyzed with automated pipelines. Total length of 1,446 genomes is 477.068 Gbp and total numbers of genes and ORFs are 6,161,842 and 8,042,823 from 185 plant genomes, respectively. The largest one is *Pinus lambertiana* (34.08 Gbp) from Gymnosperm of which average genome length is 21.08 Gbp. 237 species comprise of 4 red algae, 22 chlorophytes, one charophytes, one liverworts, two mosses, six Gymnosperm species, and 199 Angiosperm species. 26 orders of Angiosperm have sequenced genomes: Poales covers 28 species, both Lamiales and Brassicales contain 26 species. 91.43% (6,475,793) of plant ORFs have 13,321 distinct functional domains detected by InterProScan. 19,364,029 Simple Sequence Repeats (SSRs) were identified from 1,446 genomes. *Oryza brachyantha* genome have largest proportion of SSRs (4.76%) and *Arabidopsis thaliana* Castelfed-4.1 has the smallest (0.004%). Throughout these analyses, 477.068 Gbp plants genome sequence is not just collection of A, T, G, and C but new possible indicators to understand characteristics of plant genomes along with taxonomy.

P0396: General Comparative

A New Tool for Crop Breeding - rhAmp™ SNP Genotyping

Junzhou Wang[1], Krishnalekha Datta[2], Kristin Beltz[1], Daniel Tsang[2], Daniela Schrepfer[1], Yu Wang[2], Scott Rose[1], Yun Bao[2] and Caifu Chen[2], (1)Integrated DNA Technologies, Coralville, IA, (2)Integrated DNA Technologies, Redwood City, CA

Single-nucleotide polymorphisms (SNPs) have proven to be powerful genetic markers for a variety of applications in crop research and molecular breeding. Here, we report a new genotyping method called rhAmp SNP Genotyping for high-performance and cost-effective SNP genotyping in crops. This method involves cleavage and activation of blocked rhPCR primers, allele-specific PCR with universal tails, and generation of fluorescence signals using two universal reporters that specifically target allele 1 or 2 of the SNP. The activation of blocked RNA-DNA hybrid primers and target-specific cleavage at its RNA base by RNase H2 eliminate or significantly reduce primer-dimers and other non-specific interactions. Elimination of primer-dimers is the key to robust, universal, reporter-based genotyping assays. A total of 110 SNP markers from maize have been successfully genotyped at 97% design rate, 93% assay pass rate, >90% call rate, and 99.9% accuracy. Benchmarking of rhAmp technology to TaqMan® (Thermo Fisher) and KASP® (LGC) assays will also be discussed.

P0397: General Comparative

Efforts to Construct Comparative Genomics Database of Legumes and other Plant Groups

Xiyin Wang, UGA and NCST, Watkinsville, GA

Mainly due to their economic importance, genomes of 10 legumes, including soybean, wild peanuts, barrel medic, etc, have been sequenced. However, a family-level comparative genomics analysis has been unavailable. With grape and selected legume genomes as outgroups, we managed to perform a hierarchical and event-related alignment of these genomes and deconvoluted layers of homologous regions produced by ancestral polyploidizations or speciations. Consequently, we illustrated genomic fractionation characterized by wide-spread gene losses after the polyploidizations. Notably, high similarity in gene retention between recently duplicated chromosomes in soybean supported a likely autopolypoidy nature of its tetraploid ancestor. Moreover, though mostly gene losses were nearly random, largely but not fully described by geometric distribution, we showed that polyploidization contributed divergently to copy number variation of important gene families. Besides, we showed significantly divergent evolutionary levels among legumes, and by performing Ks correction, re-dated major evolutionary events during their expansion. The present effort laid a solid foundation further genomics exploration in the legume research community and beyond. We described only a tiny fraction of legume comparative genomics analysis that we performed, and more information was stored in the newly constructed Legume Comparative Genomics Research Platform (www.legumegrp.org). The present effort and recent ones in grasses have been extended into other plant families, aiming at constructing user-friendly comparative genomics platforms.

P0398: General Comparative

A Mechanism for Genome Size Reduction Following Genomic Rearrangements

Longhui Ren[1], Wei Huang[1], Ethalinda Cannon[1], David Bertioli[2] and Steven Cannon[3], (1)Iowa State University, Ames, IA, (2)University of Georgia, Athens, GA, (3)USDA–Agricultural Research Service, Corn Insects and Crop Genetics Research Unit, Ames, IA

The factors behind genome size evolution have been of great interest, considering that eukaryotic genomes vary in size by more than three orders of magnitude. Using a model of two wild peanut relatives, *Arachis duranensis* and *Arachis ipaensis*, in which one genome experienced large rearrangements, we find that the main determinant in genome size reduction is a set of inversions that occurred in *A. duranensis*, and subsequent net sequence removal in the inverted regions. We observe a general pattern in which sequence is lost more rapidly at newly distal

(telomeric) regions than it is gained at newly proximal (pericentromeric) regions – resulting in net sequence loss in the inverted regions. The major driver of this process is recombination, determined by the chromosomal location. Any type of genomic rearrangement that exposes proximal regions to higher recombination rates can cause genome size reduction by this mechanism. In comparisons between *A. duranensis* and *A. ipaensis*, we find that the inversions all occurred in *A. duranensis*. Sequence loss in those regions was primarily due to removal of transposable elements. Illegitimate recombination is likely the major mechanism responsible for the sequence removal, rather than unequal intrastrand recombination. We also measure the relative rate of genome size reduction in these two *Arachis* diploids. We also test our model in other plant species and find that it applies in all cases examined, suggesting our model is widely applicable.

P0399: General Comparative

Revisiting Pivotal-Differential Genome Evolution

Ghader Mirzaghaderi, Department of Agronomy and Plant Breeding, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran (Islamic Republic of) and **Annaliese S Mason**, Justus Liebig University, Giessen, Germany

Pivotal-differential genome relationships in the wheat (*Triticum* and *Aegilops* genera) group refer to the pattern whereby one genome in an allopolyploid is relatively conserved (pivotal) between diploid and allopolyploid wheat group species, whereas the other genome(s) is altered (differential) relative to other diploid and allopolyploid species containing this genome. This pattern was first identified based on comparative morphology (flowering spikes), whereby species could be grouped into A, D and U genome clusters. However, substantial cytogenetic evidence also supports this genome relationship, and with recent genomic advancements in wheat we suggest that it is time to interrogate this relationship further, and to extend these concepts to other plant taxa where it may be relevant. In particular, we propose that pivotal-differential genome patterning within taxa may have three possible explanations that should be tested. Firstly, variation between species sharing a differential genome may be directly inherited from variation (e.g. different progenitor cytotypes or subspecies) present within the ancestral diploid species. Secondly, variation between species may be induced as a result of the allopolyploid formation event, perhaps as a result of dominance relationships between subgenomes. Thirdly, hybridization between two allopolyploid species that share a (pivotal) genome in common but differ in their second genome may give rise to a new, rearranged (differential) genome after hybridization and genome stabilization (e.g. AABB x AACC -> AADD). Interrogation of future pan-genome data coupled with synthetic recreation of historical hybridization events is predicted to reveal the mechanisms underlying pivotal-differential genome patterns.

P0400: General Comparative

Diversity and Evolution of Plastid Transit Peptides in Higher Plants

Ryan W. Christian[1], Scott M. Schaeffer[2], Brennan L. Hyden[3], Valerie J. Lynch-Holm[4], Eric H. Roalson[4] and A. Dhingra[1], (1)Molecular Plant Sciences, Washington State University, Pullman, WA, (2)Indigo LLC, Boston, MA, (3)Washington State University, Pullman, WA, (4)School of Biological Sciences, Washington State University, Pullman, WA

Plastids take on a multitude of forms and functions in different plant organs, but temporal and interspecific variation in plastid proteomes is underexplored. Fortunately, the expanding availability of sequenced plant genomes has made it possible to do large-scale genetic comparisons between species. As a proof of concept, we examined plastid ultrastructure for three apple varieties, finding differences between apple and other species as well as between varieties, timepoints, and cell layers. This work demonstrated the scope of plastid variation within a single species, and warranted additional investigation into evolutionary change in plastid proteomes. We approached this through use of subcellular localization programs and bioinformatic clustering techniques to identify predicted shared and unique plastid-targeted peptides in 15 sequenced plant genomes. Our analysis has identified that roughly 750 proteins are shared among the plastids of all species, while individual species have between 631 to 9,185 proteins unique to their plastid proteome. Species-unique plastid-targeted proteins may exhibit novel traits even if non-plastid targeted functions are known. For high-value crops such as apple, or emerging bioenergy crops such as switchgrass, understanding how these crops differ genetically from other species presents opportunities to improve crop management for maximum yield and minimal postharvest loss. Conversely, model plants including Arabidopsis, Brachypodium, and Rice may have between 1,000 and 2,000 species-unique plastid targeted proteins each, with the ramification that experiments for these proteins will not directly translate to other plants. Further work will shed light on the evolution of plastid-nuclear trafficking to uncover how new proteins become plastid-targeted.

P0400: Other Animal Species

Convergent Amino Acid Substitutions of Avian Vocal Learning Clades

Chul Lee, IPBI, Seoul National University, Seoul, South Korea, Seoae Cho, C & K Genomics, Seoul, Korea, Republic of (South), **Heebal Kim**, Seoul National University, Seoul, Korea, Republic of (South) and Erich Jarvis, Howard Hughes Medical Institute, New York, NY

Vocal learning, the ability to imitate vocalizations based on auditory experience, is a homoplastic characteristics observed in different independent lineages of animals such as songbirds, parrots, hummingbirds and human. It has now become possible to perform proteome-wide molecular analyses across vocal learners and vocal non-learners with the recent expansion of avian genome data. Here we analyzed the whole genome of avian species that belong to one of the three vocal learning clades. We aimed to determine if behavior and neural convergence is associated with molecular convergence in polyphyletic avian vocal learners. We found molecular convergences are correlated to products of original branch lengths. In addition, we also illuminated the function of homoplastic genes specific to vocal learners was enriched for learning, and suggested a novel cAMP-based vocal learning pathway. Out of the convergent genes of vocal learning birds, *DRD5* was validated as the key candidate gene supported by multiple evidences associated with vocal learning. By applying genome editing techniques for the key gene in future, we believe phenotypic changes in transgenic birds give us insights into macro-evolution of the complex behavioral trait, vocal learning.

P0401: General Comparative

Evolution of Phytochrome-Interacting Factor (PIF) Family in Agronomically Important Angiosperms

Rafael Della Coletta, Rafael H. Gallinari, Marcelo Menossi and Mariana F. Nery, University of Campinas, Campinas, Brazil

Phytochrome-interacting factors (PIFs) are transcription factors that function as a central hub connecting many regulators of plant development such as light, hormones, temperature and the circadian clock. Despite its importance in plant growth, there are only a few species with identified PIFs, and most of the available literature in gene function is based on eight Arabidopsis genes (*PIF1* to *PIF8*). Here, we aim to identify putative PIFs in agronomically important angiosperms, understand their evolutionary history, and propose a detailed phylogenetic tree to serve as a reference for studies in this gene family. Using hidden Markov model searches (http://hmmer.org/), we identified PIFs in the genome of five monocots (Poaceae) and five dicots (Brassicaceae and Fabaceae) available at Phytozome v12.1. We also uncovered their phylogenetic relationship applying both the Maximum Likelihood and Bayesian approaches with amino acid data and bryophytes as outgroup. Our results strongly support the division of PIFs into five monophyletic groups (PIF1, PIF3/2/6, PIF4/5, PIF7 and PIF8 clades). Three of them are conserved in both monocots and dicots (PIF3/2/6, PIF4/5 and PIF8 clades), and two are exclusive to dicots (PIF1 and PIF7 clades). Lineage-specific gene duplication appears to be more common than gene loss during the evolution of PIFs, which corroborates the hypothesis that transcription factors have higher retention rates of duplicates. Our findings serve as a starting point for researchers to investigate PIF function in a broader range of plants, especially crops.

P0402: General Comparative

Snapshots of Genome Evolution in Allopolyploid Grasses

Sean Gordon, DOE Joint Genome Institute, Walnut Creek, CA

Polyploid genomes are characteristic of grasses being developed as biomass crops and many grain crops. Therefore, a deeper understanding of gene regulation and genome evolution in polyploid genomes would be useful for developing improved crop varieties for both food and fuel. Despite their economic importance and being fascinating examples of evolutionary processes in motion, it is notoriously difficult to obtain high-quality whole genome assemblies for polyploids. We are developing both computational tools and experimental systems to study allopolyploids. I will first discuss our work developing *B. hybridum* and its extant progenitor-like species (*B. distachyon* and *B. stacei*) as a simple model system to study allotetraploid genome regulation and evolution. All three species have very compact genomes, small stature and are easily grown and manipulated in the laboratory. Chromosome-level assemblies have been developed for all three species, and we have performed matched RNA-Seq and methylation experiments to reveal sequence evolution and modifications in gene regulation within the allopolyploid relative to its progenitors. We are also exploring synthetic *B. hybridum* lines obtained from crosses between the progenitor species, *B. distachyon* and *B. stacei*. In the second part of the talk, I will discuss our progress in developing computational tools to dissect complex allopolyploids *in silico*, applied to a broad range of systems, and biological insights that this enables.

P0403: General Comparative

New Insights into the Maize Pan-Genome: Incorporation of an Expanded Diversity Panel and Two Reference Genomes

Joseph Gage[1], Brieanne Vaillancourt[2], Shawn Kaeppler[3], C. Robin Buell[2] and Natalia de Leon[1], (1)University of Wisconsin-Madison, Madison, WI, (2)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, (3)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI

Recent work in plants, including maize, have shown that the genome of a species is composed of core genes that are present within all individuals and a set of dispensable genes restricted to a subset of individuals. Collectively, the core and dispensable genes compose the pan-genome. Previous work with a 503-member temperate North American diversity panel of maize identified nearly 4,000 novel transcripts not present in the reference genotype B73. To better capture the diversity available within temperate North American maize germplasm, we enlarged our diversity panel to 958-inbreds and used this expanded diversity panel to (i) better define the maize pan-genome, (ii) identify core and dispensable genome variants associated with a range of phenotypic traits, and (iii) determine how use of multiple reference genomes impacts not only defining the maize pan-genome but also in identifying sequence variation. Using whole seedling RNA-seq datasets from 958 inbred lines, we generated a pan-genome for maize using both the B73 and the PH207 reference genomes, representative genotypes of two major heterotic germplasm pools. A total of 34,447 and 39,672 novel transcripts were identified using the B73 and the PH207 reference genomes, respectively. Using sequence variants and expression potential from the 958 inbreds, we identified not only core genes but also dispensable genes associated with traits involving reproductive capacity (tassel, ear) in maize. We also show that choice of reference genotype does effect, to a small extent, sequence variants and representation of the pan-genome.

P0404: General Comparative

Transcriptome Profiling of Tissue-Specific Genes Reveals Enrichment of AP2 *cis*-Element and Phenylpropanoid Pathway in Pineapple

Neha Pandey[1,2] and Ray Ming[1,2], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China

Pineapple (*Ananas comosus* L.) is an important tropical fruit crop but lacks in the availability of extensive transcriptomic resources. Here, we analyzed the RNA-seq libraries for large-scale transcriptome profiling of pineapple plant tissue. We found 29,770 transcripts and detected expression of 20,885 genes, 16,708 of which were known genes while 4147 were hypothetical genes. The genome-wide pathway analysis reveled “phenylpropanoid biosynthesis” was enriched in roots and “carbon metabolism-CAM photosynthesis” was abundant in leaves and flowers. The flower-specific genes include SPL7, nudix hydrolase and MADS-box related transcription factors and grouped in to “regulation of circadian rhythm” and “far-red light signaling pathways”. The leaves-specific genes have phosphatidylinositol-4-phosphate 5-kinase and MYB related proteins which were involved in “ATP binding”. The root-specific genes were related to peroxidase superfamily protein, cellulose synthase and Auxin efflux protein which were involved in various response processes – oxidative stress, abscisic acid and water deprivation. Fruits-specific genes were involved in “embryo development” and “cutin biosynthetic process” and showed over-expression of Jumonji C transcription factor. The Root-, flower-, and fruit-specific genes shared AP2 motif as a conserved regulatory element. The C2H2 and bZIP motif were found in fruits specific genes and E2F3 was enriched in leaves specific genes. Moreover, in tissue-specific genes, the actual motif occurrence frequency of AP2 was five times more than the expected frequency in flower-specific genes, two times more in root-specific genes and three times more than expected frequency in fruit specific genes. The transcriptome resource will facilitate pineapple improvement via breeding and genetic engineering.

P0405: General Comparative

Ruminantia Genomes Provide Insights into Their Phylogeny and Adaptations

Lei Chen, Northwestern Polytechnical University, Xi'an, China

Ruminantia is the largest suborder of Artiodactyla, including over 200 extant species in six families (Tragulidae, Antilocapridae, Giraffidae, Antilocapridae, Moschidae, Cervidae and Bovidae). Their richness in species and wide-ranging geographical spread has arose great interests to biologists. Furthermore, the ruminants have closely relationships with humans, by serving as important livestocks like the cow, water buffalo, yak, sheep and goat. Despite the remarkable diversity and evolutionary success of the ruminants, relatively little is known about the evolutionary genomics of the group, let alone how did they evolve. Here we de novo assembled 45 ruminant genomes, covering all of the six exist families and most of the 68 exist genera of ruminantia, in an international consortium with Danish and Chinese research groups. We reconstruct the ruminant phylogeny at the resolution of whole genome and found the radiations of ruminants were closely correlated with global climate changes. We also look at a number of specific ecological and physiological adaptations such as the evolutionary innovation of the rumen and the horn. With these studies, we are hoping to bring our knowledge about the evolutionary genomics in this important animal group to a whole new level.

P0406: General Comparative

Differential Retrotransposon Activity Drives Regional Variation of DNA Turnover in Human and Mouse

Reuben M. Buckley, R. Daniel Kortschak and David L. Adelson, University of Adelaide, Adelaide, SA, Australia

Protein-coding DNA makes up less than 2% of a typical mammalian genome, however, biological roles for the remaining non-coding genomic fractions remain largely unknown. Since divergence from a common ancestor, the non-coding genomic fraction in placental mammals has undergone large amounts of lineage-specific DNA gain and loss. For example, after approximately 100 million years of DNA turnover, only 40% of the human genome shares ancestry with the mouse genome. To understand the cause and evolutionary impact of widespread DNA gain and loss, we developed a novel technique that mapped individual gain and loss events across distantly related species. By tallying the total amount of DNA turnover across genomic regions, we were able to measure associations between DNA turnover events and various regulatory/structural genomic features. Our results showed DNA loss in human and mouse mainly occurred in gene-rich open chromatin regions, whereas DNA gain was mainly driven by retrotransposition. Importantly, the level of DNA gain that associated with each retrotransposon type in both species varied significantly. This indicates differential lineage-specific activity of retrotransposons likely caused species-specific patterns of DNA gain, ultimately contributing to divergent genome architecture. To characterise potential evolutionary impacts of DNA turnover, we measured whether genes in DNA gain and loss hotspots associated with particular biological processes. Perhaps most strikingly, mouse DNA loss hotspots were enriched for genes involved in development, suggesting a potential evolutionary model where DNA loss drove evolution of morphological characteristics through turnover of gene regulatory elements.

P0407: General Comparative

A Harpin Elicitor Induces the Expression of a CC-NB-LRR Defense Signaling Gene and Others Functioning during Defense to Different Parasitic Nematodes

Vincent P. Klink[1], Weasam A. R. Aljaafri[2], Brant T. McNeece[1], Bisho R. Lawaju[2], Keshav Sharma[1], Prakash M. Niruala[1], Shankar R. Pant[1], David H. Long[3], Kathy Lawrence[4] and Gary W. Lawrence[2], (1)Department of Biological Sciences, Mississippi State University, Mississippi State, MS, (2)Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS, (3)Albaugh, LLC, Olive Branch, MS, (4)Auburn University, Auburn, AL

The bacterial effector harpin induces the transcription of the *Arabidopsis thaliana NON-RACE SPECIFIC DISEASE RESISTANCE 1*/*HARPIN INDUCED1* (*NDR1*/*HIN1*) coiled-coil nucleotide binding leucine rich repeat (CC-NB-LRR) defense signaling gene. In *Glycine max*, Gm-NDR1-1 transcripts have been detected within root cells undergoing a natural resistant reaction to parasitism by the syncytium-forming nematode *Heterodera glycines*, functioning in the defense response. Expressing Gm-NDR1-1 in *Gossypium hirsutum* leads to resistance to *Meloidogyne incognita* parasitism. In experiments presented here, the heterologous expression of Gm-NDR1-1 in *G. hirsutum* impairs *Rotylenchulus reniformis* parasitism. These results are consistent with the hypothesis that Gm-NDR1-1 expression functions broadly in generating a defense response. To examine a possible relationship with harpin, *G. max* plants topically treated with harpin result in induction of the transcription of Gm-NDR1-1. The result indicates the topical treatment of plants with harpin, itself, may lead to impaired nematode parasitism. Topical harpin treatments are shown, comparatively, to impair *G. max* parasitism by *H. glycines*, *M. incognita* and *R. reniformis* and *G. hirsutum* parasitism by *M. incognita* and *R. reniformis*. How harpin could function in defense has been examined in experiments showing it also induces transcription of *G. max* homologs of the proven defense genes *ENHANCED DISEASE SUSCEPTIBILITY1* (*EDS1*), TGA2, galactinol synthase, reticuline oxidase, xyloglucan endotransglycosylase/hydrolase, alpha soluble N-ethylmaleimide-sensitive fusion protein (a-SNAP) and serine hydroxymethyltransferase (SHMT). In contrast, other defense genes are not directly transcriptionally activated by harpin. The results indicate harpin induces pathogen associated molecular pattern (PAMP) triggered immunity (PTI) and effector-triggered immunity (ETI) defense processes in the root, activating defense to parasitic nematodes.

P0408: Canine

Thorough Evaluation of a Canine and Feline Targeted Genotyping By Sequencing Panels

Maarten de Groot, VHLGenetics, Wageningen, Netherlands

Genetic testing in cats and dogs has focused on parentage verification and monitoring genetic defects. A broad variety of technologies including TaqMan, KASP or capillary electrophoresis can be utilized to determine the markers of interest. Recent advances in targeted next generation sequencing (TGBS) technologies are creating new opportunities for service labs to expand the breadth of variants evaluated in a single, low cost test. This technological advancement led to the development of a canine and feline multiplex genetic trait panel, including the most common genetic defects amongst purebred animals and morphological traits like coat color. This poster research the use of the v1 versions of the canine and feline panel based on TGBS.

P0409: Canine

Pet Dogs, Citizen Science, and the Genomics of Behavior

Elinor K. Karlsson, Broad Institute & U Mass Med School, Cambridge, MA

Dogs are an unparalleled natural genetic model for behavioral disorders. The limited diversity in breeds, potential for admixture mapping in mixed breed dogs, and historical selection on behavioral traits makes genetic mapping approaches particularly powerful. Furthermore, as we have shown with compulsive disorders in dogs and humans, cross-species studies can elucidate the underlying neurobiology of psychiatric diseases.

We have developed a citizen-science based approach to dog behavioral genetics that includes any dog, regardless of breed ancestry, allowing us to assemble well-phenotyped cohorts of thousands or tens of thousands of dogs. In the first 2 years, "Darwin's Dogs" has enrolled 15,000 dogs, and their owners have cumulatively answered 1.5 million behavioral survey questions. Here, we describe this new approach to canine genomics. We present the first genomewide association study results using a combined single marker and admixture based approach, relying on the denser genotype information provided by the new Affymetrix dog genotyping array (400 dogs; 800,000 markers). We use owner reported phenotypes to easily identify five known size loci (IGF1, GHR, HMGA2, IGF2BP2 and SMAD2). We also find suggestive associations for complex behavioral traits, including anxiety, impulsivity and compulsive behaviors.

By engaging directly with dog owners, Darwin's Dogs is a powerful new resource for canine genomics, providing useful phenotypes and very large sample sizes. As it grows, it has the potential to facilitate recruitment for a wider range of studies, from morphology to behavior to disease.

P0410: Canine

Determining the Ancestry of American Mixed Breed Dogs

Linda Boettger, The Broad Institute, San Francisco, CA

While past genetic studies in dogs have focused almost entirely on purebreds, the majority of dogs in the United States are mixed breed, and their ancestry is not well understood. Determining ancestry in these dogs is a particular challenge due to the large number of different breeds that could contribute significant ancestry to any one dog. We conducted whole genome sequencing in 21 mixed breed dogs from diverse regions throughout the United States and determined their breed ancestry. Our study uses reference panels composed of 75 breeds, representing 92% of American purebred dogs. We implemented the *SupportMix* algorithm, which uses two machine-learning methods to assign haplotypes to specific ancestral breeds. We assessed the accuracy of the ancestry assignment through simulating breed admixture over 15 generations. On average we can determine the ancestry of haplotypes derived up to 3 generations ago with >90% accuracy and up to 12 generations ago with >75% accuracy. We found that most of the dogs in our study were mixes of at least 12 different breeds, and no dog had fewer than 6 breeds predicted with high confidence. At least 10% of the genome of each mutt derives from breeds contributing to less than 0.8% of the genome, consistent with admixture events occurring 7 generations ago or more. We conclude that American mutts are not simple mixes of a few breeds, but rather derive ancestry from numerous breeds, including admixtures from many generations ago.

P0411: Canine

Identification of Germline Variants for Body Size, Muscle and Fat Phenotypes in Domestic Dogs using 435 Canine Whole Genome Sequences.

Jocelyn Plassais[1], Brian W. Davis[2], Jaemin Kim[1], Brennan Decker[1] and Elaine A. Ostrander[1], (1)NIH-National Human Genome Research Institute, Bethesda, MD, (2)Texas A&M, College Station, TX

Domestic dogs have been subdivided into nearly 500 breeds by humans, resulting in a host of population-enriched genomic changes. As a result, modern breeds are characterized by varying working behaviors, temperament, metabolism, and aesthetic features. We have utilized our assembly of 435 canine whole genome sequences (WGS) to document over 40 million single nucleotide (SNV) and structural variants (SV) representing the largest catalog of genomic variation for a companion animal species. Using this resource, we undertook a large genome wide association study (GWAS), inclusive of 140 dogs from 99 breeds to identify loci associated with morphologic features. To validate this approach, we used GEMMA as linear-mixed model method, and first confirmed previously described loci associated with body size variation, coat color and fur phenotypes, highlighting the power of this resource. Next, we identified new quantitative trait loci associated with body size variation (weight, height), which highlighted new genes involved in bone development, fat deposition and cholesterol regulation. We next utilized the wealth of longevity data accumulated across breeds to find novel loci associated with lifespan, revealing two previously unidentified candidate loci. Finally, we have utilized the above WGS data to describe a new and complex locus on canine chromosome 12 associated with both ear morphology and height in dogs. Our results demonstrate the increasingly complex pattern of variation that controls body size and shape in dogs, while highlighting the power of this WGS catalogue to refine our understanding of canine breed traits.

P0412: Feline

Genome-Wide Evidence for Subspecies Recognition and Adaptive Divergence in the Tiger (*Panthera tigris*)

Yue-Chen Liu, Xin Sun, Lin Miao, Yan Zhuang and Shu-Jin Luo, Peking University, Beijing, China

Lack of consensus over the number of subspecies or conservation management unit in the tiger (*Panthera tigris*) has partially hindered the global effort to recover it from the brink of extinction. Today fewer than 4,000 free-ranging tigers survive in only 7% of their historical range and debates persist whether they should be considered six or two subspecies. The recent coalescence of all modern tigers to a Late Pleistocene bottleneck poses challenges to detecting subspecies-diagnostic morphological traits in the tiger and the ultimate resolution to elucidating its intraspecific evolution and taxonomy lies at the genome scale. Here based on whole genome sequencing of 32 voucher specimens we present the first genome-wide evidence that supports five statistically robust monophyletic clusters corresponding to extant subspecies, which are Sumatran (*P. t. sumatrae*), Amur (*P. t. altaica*), Indochinese (*P. t. corbetti*), Bengal (*P. t. tigris*), and Malayan (*P. t. jacksoni*) tigers, as well as one unique but tentative lineage of South China tiger (*P. t. amoyensis*) due to limited sampling. Demographic reconstructions validated a severe population decline across the whole species around 110 kya followed by serial divergent events driven by global climate fluctuations. Overall,

inter-subspecies gene flow is low corroborating these recently isolated but distinct phylogeographic units. In addition, we identified multiple genomic regions that may have formed the basis of adaptive evolution in various subspecies. These genome-wide signatures provide an explicit basis for subspecies recognition in the tiger and will help facilitate global conservation strategic planning for this charismatic flagship species.

P0413: Feline

Genomic Ancestry of the Chinese Mountain Cat (*Felis bieti*) and Domestic Cats (*F. catus*) in China

He Yu, Yueting Xing, Hao Meng, Yan Zhuang and Shu-Jin Luo, Peking University, Beijing, China

Of the 37 species of Felidae, the Chinese mountain cat (*Felis bieti*) is among the most enigmatic and only found on the Qinghai-Tibet Plateau, China. Its evolutionary ancestry and taxonomic status remain controversial, with debates centered on whether it is an independent species or conspecific with the wildcat (*F. silvestris*) and whether it may have contributed to the origin and domestication of cats (*F. catus*) in Asia. Here, based on whole genome sequencing data from Chinese mountain cats (N = 26) collected range wide and its close relatives, the Asiatic wildcat (*F. s. ornata*, N = 1) and domestic cat (N = 238) in China, we elucidated the genetic ancestry and evolutionary dynamics of this lineage and proposed a formal reclassification of the Chinese mountain cat as a subspecies of the wildcat (*F. s. bieti*). A complex hybridization scenario among wildcat lineages was also revealed, including an ancient introgression event from *F. s. ornata* into *F. s. bieti* and contemporary gene flow between *F. s. bieti* and domestic cats across but not beyond the range of *F. s. bieti*. No evidence was found that *F. s. bieti* contributed to the origin and domestication of cats in East Asia, confirming that domestic cats worldwide share a single Near Eastern origin from the African wildcat (*F. s. lybica*). Our study describes the evolutionary process and demographics of the wildcat during the Pleistocene, sheds lights on the origin of domestic cats in East Asia and provides new evidence for Felidae taxonomic reclassification.

P0414: Feline

Uncovering Causal Mutations Underlying Feline Xanthine Urolithiasis

Nicole M Tate[1], Katie M Minor[1], Leslie A. Lyons[2], Jody P Lulich[1], James R Mickelson[1] and Eva Furrow[1], (1)University of Minnesota, St. Paul, MN, (2)University of Missouri-Columbia, Columbia, MO

Hereditary xanthinuria is an autosomal recessive disease caused by mutations in xanthine dehydrogenase (XDH) or molybdenum cofactor sulfurase (MOCOS). Genetic investigations of hereditary xanthinuria in cats have previously been unsuccessful. The aim of this study was to determine the putative causal mutations underlying risk for feline xanthine urolithiasis by sequencing *XDH* and *MOCOS* in genomic DNA from affected cats. The affected cats included four Domestic Shorthairs (DSH), two of which were littermates. Sanger sequencing revealed a total of 43 homozygous variants. Multiple steps were taken to prioritize these variants and assess pathogenicity. First, the variants were genotyped in 15 control cats with no history of xanthine urolith formation (1 Ragdoll, 1 Maine Coon, 3 Siamese, 9 DSH, and 1 Domestic Medium Hair). Variants present in a homozygous state in any control were deemed likely benign. Second, the 99 Lives Cat Genome variant call file was analyzed, and variants found at an allele frequency greater than 25% were eliminated. Finally, variant effect was predicted using SnpEff and MutationTaster. Using these methods, the putative casual mutation was narrowed down to 2-4 variants per case (8 total). The variants all reside within *XDH* and include 4 missense mutations and 3 synonymous mutations predicted to alter splicing. With the exception of two shared variants in the littermate pair, the variants are unique to individual cats. In conclusion, multiple mutations with predicted pathogenicity have been identified in cats with xanthine urolithiasis. Additional prioritization steps are underway to further refine the list of causal mutations.

P0415: Feline

Evaluation of the Epilepsy Associated Genes in Familial Spontaneous Epileptic Cats Using Whole Genome Sequencing

Yoshihiko Yu[1], Daisuke Hasegawa[1], Manabu Watanabe[2], Yuji Hamamoto[1], Aki Fujiwara-Igarashi[1], Sumio Sugano[2] and Michio Fujita[1], (1)Nippon Veterinary and Life Science University, Tokyo, Japan, (2)The University of Tokyo, Tokyo, Japan

Epilepsy is a common chronic and functional cerebral disorder in both veterinary and human medicine. In 2009, a feline family lineage was identified with spontaneous recurrent seizures and a colony was established of spontaneous epileptic cat strain - 'familial spontaneous epileptic cats (FSEC)'. No feline genetic epilepsy has been identified to date. The findings of phenotypes among FSEC resembled mesial temporal lobe epilepsy, which is the most common form in humans. Therefore, FSEC is suggested as an animal model of human familial mesial temporal lobe epilepsy (FMTLE). Despite studies in multiple families with FMTLE in humans, the genetic cause has not been detected. Phenotypes of FSEC suggested the possibilities of the inheritance pattern as autosomal dominant, in addition to a complex pattern. Therefore, for the current study, 166 epilepsy-related genes that are reported in human medicine were targeted and evaluated. A cat, which shows the typical phenotypes for FSEC, was whole genome sequenced using the Ion Proton System (Thermo Fisher Scientific/Life Technologies) and aligned with the Felis_Catus-6.2 reference genome. Twenty-one of the 166 candidate genes had variants, and then Sanger sequenced in the DNA samples from four FSEC and four control cats. Four genes had variants only in the FSEC cats as either homozygous or heterozygous. However, screening in additional FSECs and control cats identified no coding variants concordant for causing epilepsy in FSEC. This data further suggests epilepsy in FSEC is not a monogenic disorder caused by a coding mutation in any of 166 known epilepsy-associated genes in human.

P0416: Feline

Genome-Wide CNV Mapping in *Felis catus* using NGS DATA

Francesca Genova[1], Maria Longeri[1], Maria C. Cozzi[1], Alessandro Bagnato[1], Maria G. Strillacci[1] and 99 Lives Consortium, (1)Università degli Studi di Milano - Dept. DiMeVet, Milano, Italy

Copy Number Variations (CNVs) represent a major source of genomic variation, affecting the phenotypic expression and onset of different diseases in humans and animals. CNVs are promising markers and their massive detection has already been performed in many domestic animals, but not yet in *Felis catus.* This work is the first CNV mapping from a large data set of whole genome sequencing (WGS) data in the domestic cat. WGS at ~30-fold genome coverage from a single ~350bp library, 150 paired-end reads out of 30 cats of nine different breeds were used (99 Lives Initiative-McDonnell Genome Institute, Washington University - St.Louis). Maverix Biomics mapped the reads on the v6.2 reference assembly. CNVs were mapped using the R cn.mops routine. Four cats were excluded as outliers. On the 26 remaining individuals a total of 164 CNVRs (106 loss, 42 gain and 16 complex regions), covering the 2.95% of the total cat genome, were detected. A

total of 89 singletons were identified and 32 CNVRs were mapped in at least five individuals. The number of CNVs in each cat ranged from 12 to 37. The GO analysis of annotated genes has been performed on Panther database. The study can be considered a starting point for genomic CNV identification in the domestic cat, aimed at disease related and breed variation analyses.

P0417: Feline

Genome-Wide Variation and Demographic History of Small Cats with a Focus on *Felis* Species

Anubhab Khan, National Center for Biological Sciences, Bangalore, India

Majority of the 38 known cat species are classified as small and they inhabit five of the seven continents. They survive in a vast range of habitats but still 12 out of the 18 threatened felids are small cats. However, there has not been enough progress in the field of small cat research as they generally get overshadowed by the charismatic big cats. Here we attempt to create a resource for small cat research especially of the genus *Felis* which has six species out of which two are classified as vulnerable by IUCN and at least one more is at risk. We collected tissue samples of four *Felis chaus* (Jungle cat) from central India and used available whole genome sequences of nine individuals from four other *Felis* species, two individuals of *Prionailurus bengalensis* and a *Otocolobus manul*. These whole genome sequences were filtered and aligned with the already published domestic cat (*Felis catus*) genome assembly. Felids are closely related species and reads from all species in our study aligned with the domestic cat genome with a rate of at least 93%. We estimated the existing genomic variation by calculating heterozygous SNP encounter rate. So far, it seems that all wild cats have more genetic variation than *Felis catus* species. This can be attributed to the inbreeding in these cats. Among the wild cats, *Felis silvestris* seems to have the highest level of genetic variation. To understand the reasons behind the distribution of genetic variation in small cats, we estimated the demographic histories of each of the species using PSMC. This method can only detect demographic changes more than 1000 generations ago. We observe that roughly all species share a parallel history in terms of population increase. The most interesting and important feature might be that all wild small cat population sizes increased exponentially around twenty thousand years ago as opposed to domestic cat and big cats which declined around this time. Another interesting feature of the demographic history is all the small cats seem to have recovered from the effects of Toba Volcano eruption which had triggered a glacial maximum leading a decline in big cat population. Thus it seems the partitioning of genetic variation has happened less than ten thousand years ago owing to anthropogenic activities?

P0418: Sheep

The Sheep Gene Expression Atlas

Emily L. Clark[1], Stephen J Bush[1], Mary B. McCulloch[1], Iseabail L. Farquhar[1], Rachel Young[1], Lucas Lefèvre[1], Clare Pridans[1], Hiu G. Tsang[1], Chunlei Wu[2], Cyrus Afrasiabi[2], Bruce Whitelaw[1], Mick Watson[1], Tom Freeman[1], Kim M. Summers[1], Alan L. Archibald[1] and David A. Hume[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)The Scripps Research Institute, La Jolla, CA

Sheep are a key source of meat, milk and fibre globally. To support functional annotation of the genome, and increase available genetic resources for sheep, we have produced a high-resolution transcriptional atlas using Texel x Scottish Blackface individuals.

RNA-Seq libraries were generated by Edinburgh Genomics (http://genomics.ed.ac.uk) from tissues and cells representing all major organ systems from adult sheep and multiple juvenile, neonatal and prenatal developmental time points. The dataset includes 367 medium depth and 74 high depth 125bp stranded Illumina RNA-Seq libraries. We have analysed the dataset using two different methods, an alignment free method, Kallisto, and a conventional alignment based HiSat2-Stringtie pipeline.

Of the protein coding genes currently annotated by Ensembl in the Oar_v3.1 reference genome, 96% are captured in the sheep atlas dataset. Using the 'guilt by association' principle based on network cluster analysis in Miru (http://kajeka.com) we have assigned meaningful gene names and putative function to hundreds of previously unannotated genes.

The sequence data and corresponding experimental metadata have been deposited in the European Nucleotide Archive (http://www.ebi.ac.uk/ena/data/view/PRJEB19199). The gene expression data is available on the BioGPS gene annotation portal (http://biogps.org/sheepatlas/) in which the pattern of expression for each gene can be visualized across tissues and cell types. Details of all tissue and cell samples can be found in the BioSamples database (https://www.ebi.ac.uk/biosamples/groups/SAMEG317052).

This comprehensive gene expression atlas for sheep provides a model transcriptome for ruminants, has the potential to inform future improvements in livestock productivity, efficiency and health, and is a valuable resource for the international Functional Annotation of Animal Genomes (FAANG) initiative.

P0419: Sheep

Rambouillet Sheep Genome and Annotation Resources

Kim C. Worley, Baylor College of Medicine, Houston, TX

We report the high quality Rambouillet sheep reference genome and initial analysis of FAANG sample RNA sequencing from the reference ewe, Benz2616.

We *de novo* assembled 200 Gb of Pacific Biosciences (PacBio) sequence with 12.6 kb N50 sub-read length with Celera Assembler and polished with Arrow. Scaffolding the contigs using Hi-C data and Phase PGA incorporated 98.1% of the assembly into 32 large scaffolds. Scaffold gaps were filled using PBJelly, misassemblies identified with misFinder and additional gap-filling completed. Error correction using Pilon and Illumina data produced the final 2.87 Gb genome. More contiguous, complete and correct than most, the contig N50 is 2.6 Mb, with half the genome in 309 contigs (longest 16.3 Mb). Most ESTs (98% of 338,551) align to the genome, 90% with nearly complete alignments, aligning over >90% of their length. Base quality is high, (error rate <1%).

FAANG assays from over 100 collected reference animal tissues including PacBio IsoSeq, Illumina RNAseq and miRNAseq, ATAC-Seq and other assays are underway, to complement the genomic PacBio, Illumina and Hi-C sequence. RNA sequence analysis of PacBio IsoSeq, Illumina RNAseq and microRNAseq (5, 9 and 20 tissues respectively) is ongoing. IsoSeq matched ~11,000 per tissue (15,888 total) of 20,921 annotated proteins. Illumina RNAseq with ~100x more reads per tissue identified ~15,500 transcripts per tissue. MicroRNA sequences analyzed using miRDeep2 identified a total of 6,523 novel miRNAs and 659 known miRNAs, with most (471) of the known and many of the novel miRNAs similar to annotated bovine miRNAs.

P0420: Sheep

ChIP-Seq Genome-Wide Identification of Regulatory Elements in Sheep Alveolar Macrophages

Alisha T. Massa[1], Michelle R. Mousel[2], Brenda M. Murdoch[3] and Stephen N. White[2], (1)Washington State University, Pullman, WA, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)University of Idaho, Moscow, ID

Lung macrophages provide first-line defense for the cell-mediated innate immune system against inhaled pathogens. Annotation of regulatory elements advances understanding of gene regulation and genetic priming of the immune system. Acetylation of lysine 27 on histone protein three (H3K27ac) is a highly dynamic histone modification, denoting active enhancer regions of the genome. Alveolar macrophages were harvested from the lungs of healthy, adult, Suffolk-cross sheep. Chromatin immunoprecipitation with high throughput sequencing (ChIP-seq) was performed for H3K27ac. Peaks denoting active enhancers were validated by comparison to mRNA-seq gene expression data publicly available in the Sheep Gene Expression Atlas. Approximately 12,000 peaks were identified in the immunoprecipitation dataset over the control with a false discovery rate below 5% and average enrichment of 18-fold. Overall, 51% of active enhancers were within or less than 500 base pairs upstream of genes. While 13% were located over 50,000 base pairs from any gene. Overall 7,964 genes were expressed, consistent with published figures of 7,000-8,000 expressed genes in macrophages. Nearly 73% of active enhancers were associated with genes in mRNA data. Overall each gene was associated with an average of 1.88 enhancers and of the validated expressed genes with multiple enhancers, each was associated with 3.46 active enhancers. Use of the newly published gene annotations from the Sheep Gene Expression Atlas allowed further identification of 859 previously unidentified genes in both the mRNA and ChIP dataset. These data provide a basis for regulatory landscape comparison among sheep cell types and for comparative regulome analysis in immune cells.

P0421: Sheep

Ovarian mRNA and miRNA Transcriptome Profiling of Domestic Sheep Breeds

Kisun Pokharel[1], Jaana Peippo[1], Mervi Honkatukia[1], Arja Seppälä[2], Johanna Rautiainen[3], Nasser Ghanem[1,4], Tuula-Marjatta Hamama[1], Mark Crowe[5], Magnus Andersson[6], Meng-Hua Li[7] and **Juha Kantanen**[1], (1)Natural Resources Institute Finland, Jokioinen, Finland, (2)Eastman Chemical Company, Oulu, Finland, (3)Lammasmaailma Oy, Ylinen, Finland, (4)Cairo University, Giza, Egypt, (5)University College Dublin, Dublin, Ireland, (6)University of Helsinki, Saarentaus, Finland, (7)Chinese Academy of Sciences, Beijing, China

The highly prolific breeds of domestic sheep (*Ovis aries*) are valuable genetic resources for global sheep industry. E.g. the native Finnsheep is well-known for its prolific traits and has been imported into 40 countries to develop new breeds and to improve fertility of local sheep breeds. Female prolificacy traits of sheep, such as the high ovulation rate and litter size, are critical factors which effect on biological and financial performance of sheep production. To improve our knowledge of the sheep prolificacy traits, we analysed mRNA and miRNA sequences of ovarian tissues from two pure breeds with large (Finnsheep) vs. small (Texel) litter sizes and their F1 crosses. Half of the ewes in each group were fed a flushing diet in order to investigate the effect of feeding on ovulation rate. We found that among the samples, 16,402 genes (60.6% known ovine genes) were expressed. We detected 79 novel miRNAs and a cluster of miRNAs on chromosome 18. The majority of the differentially expressed genes between breeds were upregulated in the low-prolific Texel, owing to the flushing diet effect, whereas a similar pattern was not detected in the Finnsheep. F1 ewes responded similarly to Finnsheep rather than exhibiting a performance intermediate between the two pure breeds. We observed in detail the variants of four major candidate genes for prolificacy (*GDF9*, *BMP15*, *BMPR1B* and *B4GALNT2*) identified previously in sheep. None of the previously studied mutations at these candidate genes were present in any of the samples. Interestingly, the mutation V371M in *GDF9* (C-T transition at 5:41841285) was present in five Finnsheep and four F1 crosses but was not present in Texel. In addition, the three genes (*CST6*, *MEPE* and *HBB*) that were differentially expressed between the group of Finnsheep and Texel ewes kept in normal diet appeared to be candidate genes of prolificacy traits and will require further validation.

P0422: Sheep

TRIM28 Regulates IGF2-H19 and Dlk1-Gtl2 Imprinting by Distinct Mechanisms during Sheep Fibroblast Proliferation

Xinhua Wang, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, China

DNA methylation is an essential epigenetic modification involved in regulating gene expression and maintaining epigenetic information across generations. However, how these marks are recognized and interpreted to activate or repress imprinted genes is not fully understood. Preliminary evidence describes the transcriptional repressor TRIM28 as a key regulator of imprinted gene expression during and after early genome-wide reprogramming. Aberrant expression of imprinted genes maybe one possible cause of incomplete epigenetic reprogramming and low efficiency in somatic cell nuclear transfer. Here, we perform a series of experiments to determine whether knock down of Trim28 alters imprinted gene expression and DMR methylation in sheep embryonic fibroblast (SEF) cells. siRNA-mediated Trim28 silencing in SEF cells resulted in significantly decreased expression of Gtl2 to 30% and increased expression of Dlk1 (~1.7-fold). Moreover, knocking down Trim28 induced DNA methylation at the IG-DMR and the Gtl2 promoter was disrupted. Here, we uncover an important role for Trim28 in the maintenance of DNA methylation at IG-DMR during replication-dependent dilution of methylated cytosine during cellular proliferation. Unlike Dlk1-Gtl2 however, knocking down Trim28 does not affect DMR methylation in the Igf2-H19 gene cluster, yet results in increased expression of Igf2 and H19. Interestingly, Peg3 expression decreased by 60% in Trim28 knockdown cells. PEG3 as a transcriptional repressor to the H19-ICR that interacts with the co-repressor protein TRIM28 through KRAB-A. Trim28 therefore appears to control the Igf2-H19 imprinted cluster indirectly via PEG3, which is distinct from its classical role in preserving DNA methylation during DNA replication. Our results therefore indicate that Trim28 regulates imprinted gene expression through at least two distinct mechanisms during cells proliferation.

P0423: Sheep

The Genomic Basis of Origin and Evolution of Headgears in Bovid and Cervid

Yu Wang, College of Animal Science and Technology, Northwest A&F University, Yang Ling, China, Wen Wang, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China; Center for Ecological and Environmental Sciences, Key

Laboratory for Space Bioscience & Biotechnology, Northwestern Polytechnical University, China and Yu Jiang, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China
Ruminants are the only group of mammals possessing osseous headgears, which is a striking morphological innovation with great advantage in evolution. Several mutants are proved as the cause mutations of hornless in cattle, sheep and goats. And breeding of hornless livestock are used to improve the production and animal welfare. However, the genetic basis of generation and regulation of headgears during mammalian evolution remains largely unknown. Taking the advantage of a large de novo genome assembly project for 50 represented ruminant species, including the headgearless Moschidae and Hydropotinae, and the available transcriptome data of horn tissues of goats and antler tissues of roe deer, we try to explore the evolution and developmental regulation of headgear. We identified positive selection genes and gene family expansion and contraction, highly specific divergent elements to depict the evolution of headgear. In addition, we explored the mechanisms of genetic mutation of headgear, including hornless and polyceraty. Our study shed new light on genetic and regulated mechanisms underline ruminant headgears and contributes to product hornless ruminant livestock.

P0424: Sheep

Differential Gene Expression Analysis of Skin Tissue in Follicle Anagen between Chinese Merino and Kazakh Sheep

Min Shen, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, China
Using Illumina HiSeqTM2000 high-throughput platform and bioinformatics technologies, we carried out whole-genome RNA sequencing on the skin tissues in anagen of Chinese Merino (superfine wool type) and Kazakh sheep (coarse wool breed), analyzed and compared the potential transcriptome differences of skin tissue during follicle anagen phase between different sheep populations.

P0425: Sheep

Characterizing Allelic Variation in the Recombination Hotspot Mediator Gene PRDM9 in U.S. Sheep

Kimberly M. Davenport[1], Theodore S. Kalbfleisch[2], Stephanie McKay[3], Michael P. Heaton[4] and Brenda M. Murdoch[1], (1)University of Idaho, Moscow, ID, (2)University of Louisville, Louisville, KY, (3)University of Vermont, Burlington, VT, (4)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE
Meiotic recombination is an important process during gametogenesis that ensures proper chromosome segregation and contributes to genetic variation. It is clear from previous studies that recombination events are not entirely random, and exhibit distinct location preferences with high activity, termed “hotspots.” The *PRDM9* gene encodes a zinc finger protein that mediates hotspot usage and locations in mice, humans, and presumably sheep. In addition, incompatibility between *PRDM9* alleles is known to cause sterility in male mice. In this study, we characterized *PRDM9* alleles and identified two (C and D) not previously reported in U.S. sheep. These alleles were identified by the size of the zinc finger region, and were similar to those in populations abroad. The D allele (9 zinc fingers in length) was the most common, and the C allele (8 zinc fingers in length) was only observed in a heterozygous CD genotype. To further characterize *PRDM9* in U.S. sheep, we identified haplotypes in the zinc finger region using the USMARC Sheep Diversity Panel v2.4 which differed from those previously reported. Characterizing *PRDM9* will allow us to better discern the use and location preferences of recombination hotspots in the genomes of U.S. sheep populations. Further, defining *PRDM9* alleles in U.S. sheep will enhance our ability to understand whether they significantly affect reproduction in males or the accuracy of genetic predictions.

P0426: Sheep

Genome-Wide Association with Monocyte Count in Domestic Sheep

Ryan D. Oliveira[1], Michelle R. Mousel[2], Michael V. Gonzalez[3], J. Bret Taylor[4], Donald P. Knowles[3] and Stephen N. White[2], (1)Washington State University College of Veterinary Medicine, Department of Veterinary Microbiology and Pathology, Pullman, WA, (2)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (3)Washington State University, Pullman, WA, (4)USDA, ARS, Range Sheep Production Efficiency Research, Dubois, ID
Monocytes and macrophages are an essential line of phagocytic and antigen presenting immune cells that play a direct role in innate immunity that can stimulate adaptive immune responses. These cells have been implicated in susceptibility to many infectious diseases of sheep, including small ruminant lentivirus. Therefore, counts of circulating blood monocytes per unit blood volume were collected by automated high throughput methods for 514 Rambouillet, Polypay, and Columbia sheep (*Ovis aries*). A genome-wide association study (GWAS) investigating monocyte counts was performed with the OvineSNP50 genotyping array. A general linear model accounted for breed, multidimensional scaling cluster, and age, in addition to the SNP of interest. A genome-wide significant peak was observed on ovine chromosome 9 and genome-wide suggestive peaks on chromosomes 1, 3, 4, 5, 11, 15, and 23. Nearby genes included hematopoietic and developmental genes. Further investigation is underway to refine the positions and functions of these loci. Differing monocyte numbers may affect susceptibility to infectious disease, and investigating the genetic basis by which monocytes are produced and maintained may add value to fundamental monocyte and macrophage biology.

P0427: Sheep

Tenascin-Xb (TNXB) Amino Acid Substitution E2004G Is Associated with Mature Weight and Milk Score in American Rambouillet, Targhee, Polypay, and Suffolk Sheep

Mehmet Ulas Cinar[1,2], Michelle R. Mousel[2,3], Maria K. Herndon[2], J. Bret Taylor[4] and **Stephen N. White**[2,3], (1)Erciyes University, Kayseri, Turkey, (2)Washington State University, Pullman, WA, (3)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (4)USDA, ARS, Range Sheep Production Efficiency Research, Dubois, ID
Sheep are economically important with worldwide distribution of approximately 1 billion animals. Growth-associated genetic variants, such as a *TNXB* charged amino acid substitution E2004G and a silent *DGAT1* single nucleotide polymorphism (rs409119650), could improve profitability of sheep production. However, both were identified in single research reports utilizing small groups of sheep. We evaluated 896 U.S. sheep to investigate association of *TNXB* E2004G and *DGAT1* rs409119650 with growth and lifetime production. For *TNXB* E2004G, glutamic acid homozygotes had greater live body weights in spring and fall at ages 3 and 4 (all $P \leq 0.05$) and greater milk scores at ages 3 and 4

($P < 0.05$). The milk scoring system used was economically relevant since it had previously been shown to correlate with lamb growth. Thus, these data provide the first report of an association with an economically important milk measure for *TNXB* E2004G, as well as confirming a prior association of the ancestral glutamic acid allele with increased growth. While *DGAT1* rs409119650 was associated with increased 4-year-old live body weight and lifetime greasy fleece weight ($P < 0.05$), these results should be interpreted with caution given the low observed minor allele frequency (4.9%) and additional datasets are needed to fully evaluate utility of *DGAT1* rs409119650. Overall, the observed *TNXB* E2004G association with increased live weight in divergent breeds and multiple countries may suggest a useful role in selective breeding programs, but careful optimization of breeding strategies may be valuable to balance growth rates and mature body size appropriately.

P0428: Sheep

Genomic Regions Associated with Entropion Affecting One or Both Eyes of Domestic Sheep

Michelle R. Mousel[1,2], Stephen N. White[1,2] and Maria K. Herndon[2], (1)USDA, ARS, Animal Disease Research Unit, Pullman, WA, (2)Washington State University, Pullman, WA

Entropion is an inward rolling of the eyelid allowing eyelashes to contact the cornea, potentially causing abrasions which may lead to infections and blindness if not treated. In domestic sheep, entropion usually occurs with the lower eyelid and is a congenital defect. Entropion affects domestic sheep globally at a greater frequency than other afflicted mammals, was found to be heritable (0.08-0.21), and is thought to be recessive in inheritance. Identification of genomic regions or genes associated with entropion could lead to the development of genetic marker(s) to reduce entropion through selective breeding. Therefore, a genome-wide association scan was conducted with 489 Columbia, Polypay, and Rambouillet sheep genotyped using the Illumina OvineSNP600 BeadChip. Entropion status as none, 1, or both eyes was recorded within 48 hours of birth and corrected if present. There were 43 sheep categorized as having entropion. Data was analyzed using a mixed model with EMMAX that accounted for relatedness and breed. Heritability was estimated to be 0.82. Nine genome-wide significant ($P<1x10^{-8}$) SNPs were identified on chromosomes 1, 2, 8, 17, 18, 19, and X as well as twelve SNPs that were genome-wide suggestive ($P<1x10^{-7}$) on chromosomes 1, 2, 6, 8, 9, 15, 20, and X. As a binomial trait, previous research found locations on ovine chromosomes 2, 3, 6, 15, and 16 to be associated with entropion. We are working to narrow the range of these associated regions and identify the underlying casual mutations to improve sheep health and elucidate entropion in other mammals.

P0429: Sheep

Signatures of Selection in Ethiopian Sheep and Cattle Populations Adapted to Diverse Environments

Kwan-Suk Kim, College of Agriculture, Life & Environmental Sciences, Chungbuk National University, Cheongju, South Korea, Zewdu Edea Bedada, Chungbuk National University, Chungcheongbuk-do, Korea, Republic of (South), Dessie Tadelle, International Livestock Research Institute, Addis Ababa, Ethiopia and Hailu Dadi, Addis Ababa Science and Technology University, Addis Ababa, Ethiopia

Ethiopia is a gateway for multiple livestock introductions into Africa and endowed with rich sheep and cattle genetic resources. Some sheep breeds such as the Blackhead Somali and the Afar sheep; and cattle breed such as Begait, Borana and Ogaden are well adapted to arid lowland areas, whereas other sheep breeds (Menz and Arsi-Bale) and cattle (Arsi) thrive under cool high-altitude areas. These indigenous populations can serve as a model to investigate signatures of selection for environmental adaptation. However, little research has been conducted to map genes or genomic regions linked to ecological adaptation within and across species. To identify selection signatures in sheep populations adapted to highly contrasted environments, 72 individuals representing high-altitude (2000-3000 m a.s.l) and low-altitude (500-1000 m. a.s.l) populations were genotyped with the Illumina Ovine Infinium HD SNP BeadChip (600K). In cattle, 125 animals sampled from high-and-low altitude environments were genotyped with 80K BeadChip mainly derived from *Bos indicus* breeds. F_{ST} and run of autozygosity were calculated across sliding windows between the low-and- high altitude populations. The detected candidate genes are involved in high-altitude adaptation (*DNAH9, PRKAA1,* and *MLLT10*), hair type and length (*FGF5*), pigmentation (*MITF, KITLG,* and *MC1R)* and fat deposition (*MC2R* and *MC5R*). Interestingly, we detected common genes including *PSPC1, MPHOSPH8, PARP4,* and *RNF17* subjected to selection in cattle and sheep on Bovine 12 and Ovine 10, respectively. The results of this study could be used to further dissect the genetic basis of adaptation to extreme environments in ruminant livestock species.

P0430: Sheep

Effects of Maternal Diets on Fetal Genomic Imprinting in Sheep

Jingyue (Ellie) Duan[1], Mingyuan Zhang[2], Kaleigh Flock[1], Sahar Al Seesi[3], Ion Mandoiu[3], Amanda K Jones[1], Sambhu M Pillai[1], Maria L Hoffman[1], Hesheng Jiang[2], Sarah A Reed[1], Kristen E Govoni[1], Steve A Zinn[1], Zongliang Jiang[1,4] and Xiuchun (Cindy) Tian[1], (1)Department of Animal Science, University of Connecticut, Storrs, CT, (2)College of Animal Science and Technology, Nanning, China, (3)Department of Computer Science, University of Connecticut, Storrs, CT, (4)School of Animal Science, Louisiana State University, Baton Rouge, LA

Genomic imprinting is an epigenetics phenomenon that causes differential allelic gene expression based on parental origin. To date, 255 imprinted genes have been identified or predicted in all mammals combined. However, imprinting study in sheep lags behind, as only 21 imprinted genes have been described. The current study used DNA/RNA throughput sequencing to identify monoallelically expressed and imprinted genes in day 135 sheep fetal organs, and to access the influence of maternal nutrition on imprinting. We solved several technical challenges in NGS data analysis pipeline including alignment bias of RNA sequencing reads and filtering potential false positives. We identified 80 monoallelically expressed gene and 18 imprinted genes, five of which were previous known imprinted in sheep, and thirteen were known imprinted in other species. Sanger sequencing confirmed four new sheep imprinted genes *INPP5F, PLAGL1, CASD1* and *PPP1R9A*. Among the thirteen new imprinted genes, five located in the sheep known imprinting clusters of *MEST* domain on chromosome 4, *DLK1/GTL2* domain on chromosome 18 and *KCNQ1* domain on chromosome 21, three were in a novel sheep imprinted cluster on chromosome 4 known in other species as *PEG10/SGCE*. Additionally, we found *PHLDA2, SLC22A18, DIRAS3*, and *IGF2* differentially expressed, but no allele expression reversal or loss of imprinting among three maternal nutritional groups. Our results expand the imprinted gene list to 34 in sheep and demonstrate the influence of maternal diet on fetal imprinting in sheep under the conditions studied.

P0431: Sheep

Transcriptomic Analyses Reveals Peroxisome Proliferator Activated-Receptor Signaling Pathway As a Potential Regulator Mechanism in Gastrointestinal Parasite Infections in Sheep

Pablo Peraza, Instituto Nacional de Investigación Agropecuaria, Las Brujas, Uruguay, Gonzalo Rincon, Zoetis, Kalamazoo, MI, Gabriel Ciappesoni, National Institute of Agriculture Research, Rincon del Colorado, Canelones, Uruguay, Alma Islas-Trejo, University of California-Davis, Davis, CA, Juan F. Medrano, University of California, Davis, CA and José Sotelo-Silveira, Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay

The underlying mechanisms of gastrointestinal parasite (GIP) resistance in sheep remains unclear. Research divergent selection flocks containing lines that are resistant (R) or susceptible (S) to gastrointestinal nematodes, provide a unique resource for studying an effective nematode response in animals. In this work, we compared gene expression levels from different tissues (abomasum, duodenum, jejunum, ileum and lymph nodes) in resistant and susceptible animals. We observed that peroxisome proliferator activated receptor (PPAR) appears as a potential regulator of this resistance when analyzing all tissue samples. PPAR plays an essential role in lipid metabolism and has a major role as a transcription factor regulating the expression of several genes involved in cell differentiation and development. Duodenum and Jejunum tissues showed a high association with expression of PPAR. Both tissues are directly related with the parasitic site of infestation. In addition, FXR/RXR receptor activation was also regulated as a top canonical pathway modulating gene expression. We also detected several genes as mucin or intelectin (MUC / ITLN) involved in immune and damage responses. These mechanisms could facilitate an enhanced selection for parasite resistant sheep through marker/gene-assisted selection or future targeted gene manipulations to control GIP resistance.

P0432: Sheep

TRIM28 regulates IGF2-H19 and Dlk1-Gtl2 imprinting by distinct mechanisms during sheep fibroblast proliferation

Limin Wang, XinJiang Academy of Agricultural and Reclamation Science, Shihezi, China

Fat-rumped Altay sheep provide an ideal model to study fat mobilization and deposition. Altay sheep display the unique ability to rapidly mobilize fat to maintain normal metabolism that facilitates their survival in long-term harsh winter conditions, but the underlying physiological, biochemical and molecular mechanisms remain unclear. In this study, we monitored the levels of TG, TCH, FFA and two adipocyte-specific hormones in Altay sheep at three different stages over a one-year period, examined the morphology of trump-fat adipocytes in different sheep breeds and at distinct stages of the life cycle, and established a persistent starvation model to imitate the two extreme states of rump fat deposition and mobilization. Simultaneously, expression patterns of lipid metabolism-related genes in rump fat, selected using iTRAQ, were assessed before and after the starvation test. TG, TCH and FFA levels were not significantly different between the groups ($P \geq 0.05$). The rump-fat sectioning data showed a positive correlation of cell size with fat deposition ability. iTRAQ proteomic analysis facilitated the detection of ~1610 proteins in Altay rump fat, among which 112 displayed significant differences in expression between the two extreme states. On this basis, we observed increased secretion of LEP and ADIPOQ and simultaneous activation of key fat mobilization and fatty acid oxidation signal pathways, such as AMPK and PPAR, under persistent starvation and extreme cold conditions. We hypothesize that upregulation of *RETN*, *HSP72* and *CFD* enhances insulin resistance to promote lipolysis, downregulation of *CIDEC* inhibits lipid droplet fusion, and upregulated *HSP72* and *Apo-AI* activate body stress and the anti-inflammatory mechanism. The synergistic actions of the above hormones, genes and signal pathways provide a molecular basis and material guarantee for improving adaptability of Altay sheep to extreme environments. Our collective findings serve as a reference for elucidating the complex molecular mechanisms of rump fat deposition and mobilization.

Keywords Sheep, rump fat, Protein expression, LC-MSMS

P0433: Goats

Frequencies of Y-Chromosome Haplotypes in Goat Breeds from Spain

Marcel Amills, Center for Research in Agricultural Genomics, BARCELONA, Spain and Oriol Vidal, Universitat de Girona, Girona, Spain

The main goal of the current work was to characterize Y-chromosome variation in Spanish goats by genotyping seven single nucleotide polymorphisms (SNPs), mapping to the SRY, ZFY, AMELY and DDX3Y Y-linked genes in 78 bucks from seven Spanish breeds. In general, the amount of diversity was quite low, with the majority (98%) of Spanish goats carrying the Y2 haplotype. A comparative analysis with paternal variation in goats from Central Europe revealed a high Y-chromosome differentiation between those and the Spanish ones. This result might be explained by the post-domestication dispersal of different genetic stocks across Europe as well as by historical gene flow between Spanish and Northern African goats.

P0434: Goats

Selection Signatures and Genetic Diversity in Specialized and Locally Adapted Goat Breeds in Americas

Tiago do Prado Paim[1,2], El Hamidi Hay[3], Concepta McManus[4], Harvey Blackburn[5], Danielle A. Faria[6], María Rosa Lanari[7], Laura Chaverri Esquivel[8], María Isabel Camacho Cascante[8], Argerie Cruz Méndez[9], Carlos Alberto Mezzadra[10], Arthur da Silva Mariante[11] and Samuel Rezende Paiva[11], (1)Universidade de Brasília / NAGP-ARS-USDA, Fort Collins, CO, (2)Instituto Federal de Educação, Ciência e Tecnologia Goiano, Iporá, Brazil, (3)USDA Agricultural Research Service, Miles City, MT, (4)Universidade de Brasília, Brasília, Brazil, (5)National Center for Genetic Resources Preservation, USDA, Fort Collins, CO, (6)Institute of Biological Sciences (IB) - Universidade de Brasília (UnB), Brasília, Brazil, (7)Instituto Nacional de Tecnologia Agropecuaria (INTA), San Carlos de Bariloche Río Negro, Argentina, (8)Universidad Nacional de Costa Rica, Heredia, Costa Rica, (9)Instituto de Innovación y Transferencia en Tecnologia Agropecuaria (INTA), San José, Costa Rica, (10)Instituto Nacional de Tecnologia Agropecuaria (INTA), Balcarce Buenos Aires, Argentina, (11)Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil

Signatures of selection are useful to understand association between genomic regions and key phenotypic traits. Our objective was to detect selection signatures among 12 goat breeds from five American countries. Samples and SNPs with call rate < 0.90 were removed, as well as monomorphic, unmapped and allosome SNPs. Samples with genomic relationship higher than 0.25 were removed, leaving 246 animals and 48,442 SNPs. Genetic structure analysis showed 5 distinct groups (Angora [US, Argentina and South Africa]; Boer [US]; Milk – LaMancha [US] and Saanen [Costa Rica]; Brazilian – Moxoto and Caninde; Spanish [US] and Argentinean locally adapted breeds). Those groups were considered to perform pairwise Fst analysis. Smoothed Fst greater than three standard deviations above mean were considered as significant. Gene annotation was with NCBI Genome Data Viewer and gene ontology with PANTHER. There were eleven regions among all pairwise comparisons. The Argentinean breeds showed two regions with 22 genes, which were mainly related to immunological systems and biological regulation. The Angora and Milk breeds showed four regions, containing 82 and 27 genes for each group, respectively. The genes significant for Angora were related to response to stimulus and localization. Chromosome 11 had one region significant for milk breeds with *BRISC* and *BRCA1* which are associated with several types of breast cancer in humans. The selective sweeps found among breeds utilized for the same function (e.g., meat, dairy, fiber) despite broad geographic differences suggest potential for new approaches in designing phenotypic experimentation and conservation among countries.

P0435: Goats

The Genetic Diversity in Body Size of African Goats

John C. Nystrom[1], Mary Beth Hannon[2], Elizabeth A. Staiger[1], M.J. Woodward-Greene[3], Johann Soelkner[4], Benjamin D. Rosen[5], Curtis P. VanTassell[6], Tad S. Sonstegard[7], Heather J Huson[1] and African Goat Improvement Network (AGIN), (1)Cornell University, Ithaca, NY, (2)Cornell University, Department of Animal Science, Ithaca, NY, (3)USDA-ARS, Beltsville, MD, (4)University of Natural Resources and Life Sciences, Vienna, Vienna, Austria, (5)ARS, USDA, Beltsville, MD, (6)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (7)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

Small holder farmers in rural Africa rely heavily on goat production because of their hardiness and ability to survive in sparsely vegetated areas. Diverse populations of goats adapted to unique ecosystems and management styles provide meat, milk, and income to families. The African Goat Improvement Network (AGIN) is a USAID Feed the Future collaboration with over 20 countries representing livestock and agricultural research institutions and universities, a majority of them African. One aim of AGIN is to phenotypically and genetically characterize body size, a particularly important attribute directly affecting income potential in the meat market. Multiple genome-wide association studies (GWAS) were conducted exploring the genetic regulation of body size in African goats.

The project utilized 968 goats from 13 African countries reflecting 58 breeds/populations. Goats were genotyped on the Illumina Caprine 50K beadchip using 51,005 SNPs after quality assessment. Data collected included 5 body size measures, sex, breed, owner, and location. Principal Component Analysis of body measurements produced PC1 reflecting overall body size. The effect of weight was explored to identify its impact for GWAS. PC1 was used as the quantitative variable in an optimized mixed model linear analysis with covariates of sex and breed under a recessive model. The GWAS on body size highlighted 7 SNPs spanning chromosomes 1, 5, 11, 18, and 26 with weight showing no significant impact on outcomes. Positional candidate genes are currently being investigated for sequence variation to provide insight on the genetic regulation of body size in African goats.

P0436: Goats

Identification of Selection Signals by Large-Scale Whole-Genome Resequencing of Cashmere Goats

Rui Su[1], Xiaokai Li[2] and Jinquan LI[1], (1)Inner Mongolia Agricultural University China, HOHHOT, China, (2)College of Animal Science, Inner Mongolia Agricultural University, Hohhot, China

Inner Mongolia and Liaoning cashmere goats are two outstanding Chinese multipurpose breeds that adapt well to the semi-arid temperate grassland. These two breeds are characterized by their soft cashmere fibers, thus making them great models to identify genomic regions that are associated with cashmere fiber traits. Whole-genome sequencing of 70 cashmere goats produced more than 5.52 million single-nucleotide polymorphisms and 710,600 short insertions and deletions. Further analysis of these genetic variants showed some population-specific molecular markers for the two cashmere goat breeds that are otherwise phenotypically similar. By analyzing F_{ST} and q_p outlier values, we identified 135 genomic regions that were associated with cashmere fiber traits within the cashmere goat populations. These loci contained genes, such as *FGF5, SGK3*, *OXTR*, and *ROCK1*, which are involved in the production of cashmere fiber. Gene ontology enrichment analysis of identified short insertions and deletions also showed enrichment in keratinocyte differentiation and epidermal cell differentiation. These findings demonstrate that this genomic resource will facilitate the breeding and genetic basis of economically traits of cashmere goat and other *Capra* species in future.

P0437: Goats

Transcriptomics of Innate Immune Response in the Domestic Goat and Comparative Analysis with Sheep

Charity Muriuki[1], Mike Coffey[2], Morris Agaba[3], David A. Hume[1,4] and Emily L. Clark[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)Scotland Rural College (SRUC), Edinburgh, United Kingdom, (3)Sarissa Ltd, Private box, Kampala, Uganda, (4)Mater Research, University of Queensland, Translational Research Institute, Woolloongabba, Australia

Despite great genetic similarity, ruminants vary in their susceptibility to similar pathogens but the underlying molecular mechanisms remain largely unknown. To elucidate the molecular basis of variation in disease response, we generated a gene expression atlas of the domestic goat from a subset of tissue and cell types and compared it to sheep and other ruminants. We produced 54 TruSeq 75bp paired-end libraries on the Illumina platform at 25M reads per sample including six libraries from bone-marrow derived macrophages stimulated with lipopolysaccharide to mimic pathogen challenge by gram-negative bacteria. Transcripts were quantified using Kallisto to generate gene expression estimates as transcripts per million (TPM), capturing nearly 90% of the annotated protein-coding genes in the goat reference transcriptome. We visualize the data as gene-gene network graphs in Miru (http://kajeka.com) to investigate tissue-specific gene expression profiles and assign function to

unannotated genes through principle of guilt-by-association. Additionally, using percentage similarity of protein-identity between goat and sheep, we are running a comparative transcriptomic analysis with the recently released high-resolution gene expression atlas of the domestic sheep. We identify numerous, previously unannotated genes involved in innate immune response in goat and provide useful data to improve annotation of the goat reference genome (ARS1). The sample metadata have been loaded into BioSamples and sequence data submitted to the ENA available for use by the wider research community. This project aids the understanding of molecular mechanisms controlling variation in disease susceptibility across ruminants and provides a valuable resource for the study of ruminant functional genomics.

P0438: Swine

A Comprehensive Map of cis Regulatory Elements and 3D Structure of the Pig Genome

Yu Luan[1], Yunxia Zhao[1], Ye Hou[1], Yueyuan Xu[1], Huanhuan Zhou[1], Hongbo Yang[2], Feng Yue[2], Xinyun Li[1] and Shuhong Zhao[1], (1)Key Lab of Agricultural Animal Genetics, Breeding, and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan, China, (2)Department of Biochemistry and Molecular Biology, College of Medicine, The Pennsylvania State University, Hershey, PA

Pig is an important livestock species for meat supply. Moreover, as ~80% of the protein-coding genes are conserved between human and pig, pig is also an ideal model for biomedical studies.However, the functional annotation of pig genome is largely behind comparing with human and mouse. Here we took a similar approach adopted by the encode and roadmap epigenomics projects, and performed RNAseq and ChIPseq for H3K27ac and H3K4me3 histone markers to generate a comprehensive map of transcriptomes and regulatory element in a variety of pig tissues including muscle, fat, liver, heart, and spleen from Large White, Duroc, Meishan and Enshi Black pigs. We obtained over 10,000 cis regulatory elements in the pig genome, the most comprehensive functional annotation effort made so far in the pig. By comparing the data generated by the ENCODE and Roadmap Epigenomics projects, we also defined a set of functionally conserved and species-specific regulatory sequences among pig, mouse and human.To further explore the three dimension(3D) structure of pig genome, we performed (high-throughput chromosome conformation capture (Hi-C)) experiment using pig muscle and liver tissues. Our HiC matrix showed that topologically-associating domains were also conserved among pig, mouse, and human. Muscle and liver specific enhancer and promoter interactions predicted from CHIP-seq data are further validated by muscle Hi-C matrix.
In summarys, we generated a great genomics / epigenomic resource for the functional annotation and the 3D structure of the pig genome and further expanded the value of pig

P0439: Swine

Additional Annotation of the Pig Transcriptome using Integrated Iso-Seq and Illumina RNA-Seq Analysis

Hamid Beiki[1], Haibo Liu[2], Jianzhen Huang[3], Nancy Manchanda[4], Dan Nonneman[5], Timothy P.L. Smith[5] and Christopher K. Tuggle[1], (1)Iowa State University, Ames, IA, (2)Molecular, Cell and Cancer Biology, UMass Medical School, Worcester, MA, (3)Jiangxi Agricultural University, Nanchang, China, (4)Genetics, Development and Cell Biology Department, Iowa State University, Ames, IA, (5)USDA, ARS, USMARC, Clay Center, NE

Alternative splicing is a well-known phenomenon that dramatically increases eukaryotic transcriptome diversity. The extent of mRNA isoform diversity among porcine tissues was assessed using Pacific Biosciences single-molecule long-read isoform sequencing (Iso-Seq) and Illumina short read sequencing (RNA-seq) from a single individual White cross-bred pig. Isoseq data for nine tissues (brain, hypothalamus, liver, muscle, thymus, pituitary, small intestine, spleen and diaphragm) was error-corrected using RNA-seq data from the same RNA samples. Alignment of RNA-seq data to the current Ssc11.1 build (Ensembl release 90) for all 9 tissues revealed 401 tissue specific (TS) genes (50-fold higher FPKM level in one tissue compared with all others) and 8,309 housekeeping genes (FPKM≥1 in all tissues). Interestingly, 262 TS genes had no Gene Ontology annotation. Integration of IsoSeq and RNAseq data in liver and brain tissues identified 17,086 expressed (RNAseq FPKM≥1) novel isoforms (isoform that have at least one novel splice junction with an annotated transcript). Many expressed isoforms were detected within annotated intergenic (1,271) or intronic (801) regions. Many of these novel genes were validated using H3K36me3 (gene body mark) and H3K4me3 (promoter mark) ChIP-seq in liver. Analyses are ongoing for the other seven tissues, as well as tissue specific alternative splicing and long non-coding RNAs across all tissues. In summary, the results of this study can improve the incompletely annotated pig genome through the addition of alternative splicing complexity and identification of new features that are not included in the current pig genome annotation. The USDA is an equal opportunity provider and employer.

P0440: Swine

Genomic Co-Localization of microRNA Expression QTL Target Genes with Phenotypic QTL in the Michigan State University Duroc x Pietrain Pig Resource Population

Kaitlyn Daza, Deborah Velez-Irizarry, Sebastian Casiro, Juan P. Steibel, Nancy E. Raney, Ronald O. Bates and Catherine W. Ernst, Department of Animal Science, Michigan State University, East Lansing, MI

MicroRNAs (miRNAs) are a class of non-coding RNAs known to post-transcriptionally regulate gene expression through binding with target mRNAs, ultimately affecting a multitude of biological processes and phenotypes. Combining miRNA and gene expression profiles with genotypic and phenotypic data from the same animals enables the elucidation of regulation of complex traits important to the pig industry. The objective was to identify genomic co-localization events of miRNA expression Quantitative Trait Loci (miR-eQTL) miRNA target genes from *Longissimus dorsi* muscle samples with previously-identified phenotypic QTL (pQTL) in the MSU Duroc x Pietrain population. Animals were previously characterized for over 60 phenotypes and genotyped with Illumina PorcineSNP60 BeadChips. In total, 295 mature miRNA expression profiles were included in a GBLUP-based GWA analysis. Target genes for 15 miR-eQTL miRNAs were identified using TargetScan and filtered based on transcript abundance data from the same samples. Target genes negatively correlated (FDR≤0.05) with their associated miRNA's expression were co-localized with pQTL, yielding three miR-eQTL miRNAs with 29 total target genes overlapping pQTL across seven chromosomes. One miR-140-5p target, *RRP36*, co-localized with pQTL for dressing percentage (SSC7), while three targets of miR-6782-3p co-localized with pQTL for number of ribs (SSC7). Targets of miR-874 co-localized with a large pQTL on SSC15 for meat quality traits including juiciness, tenderness, Warner-Bratzler shear force, protein content, pH at 24 h, drip loss, and cook yield. Continuing to

investigate miR-eQTL and their effects on downstream phenotypes will contribute to deciphering mechanisms controlling complex pig production traits.

P0441: Swine

Integrative Alignments of DNA Elements for Transcriptional Regulation in Swine Epigenome

Jianhua Cao, Huazhong Agricultural University, Wuhan, China

The state of DNA elements is important for gene transcription and regulation. However, the epigenomic characterizations still remain largely unknown in the pig genome, and the relationship between epigenetic modification and gene expression is thus much less investigated on whole genome wide scale. Here, we applied ChIP-Seq combined with RNA-Seq strategy to comprehensively align chromatin states and gene expression on 8 epigenomic markers (H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3, H3K36me3, RNAPII and CTCF) in 3 types of porcine cells (macrophage cell line 3D4/21, kidney cell line PK-15 and primary alveolar macrophage PAM). Based on the histone modifications, we established 19 chromatin states which represented 8 active and 7 repressed states in swine epigenomes. Also, we identified the common regions with similar chromatin states of gene transcriptions in these cells, and whereas found cell type specific modification patterns in swine genome. Of particular, we introduced RNAPII and CTCF factors on purpose to pre-define chromatin conformation associated information to assist research on porcine three-dimensional (3D) chromatin organizations in the future. The relationship between gene transcription and DNA proximal modifications demonstrated that transcriptional regulations largely depending on the chromatin states of cis-elements in swine epigenome. Our findings in this study thus provide new clues and primary information to explore the encyclopedia of DNA elements (ENCODE) and are also likely to shed light on elucidating chromatin organization and dynamics underlying the epigenomic mechanism in swine.

P0442: Swine

Transcriptional and Genome-Wide Methylation Profiling in Fetal Pig Skeletal Muscle

Ryan J Corbett[1], Nancy E. Raney[2] and Catherine W. Ernst[2], (1)Genetics Graduate Program, Michigan State University, East Lansing, MI, (2)Department of Animal Science, Michigan State University, East Lansing, MI

Development, growth, and function of skeletal muscle are dynamic processes. Formation of skeletal muscle fibers occurs during pig fetal development in two waves: primary fibers form de novo (30-60dg) and secondary fibers form around primary fibers (54-90dg). Changing biological demands of tissues during development require coordinated transcriptional regulation; DNA methylation is one mechanism impacting this process. We used *Longissimus dorsi* muscle samples from pigs at 41dg and 70dg (n=3 per stage) to determine changes in CpG methylation status (whole-genome bisulfite sequencing, WGBS), and transcript abundance (RNA-seq and miRNA-seq) between stages. Sequencing was performed on the Illumina HiSeq 4000 platform. Bismark was used to align WGBS reads and obtain methylation rates at each CpG, and differential methylation (DM) analysis was performed using the methylKit R package. RNA-seq and miRNA-seq reads were mapped to the *S. scrofa* reference genome (v.11.1) using TopHat2 and miRDeep2, respectively; differential expression (DE) analysis was performed using DESeq2. Significant negative and positive correlations were observed between gene expression and methylation in promoters ($r = -0.26$, $p<0.001$) and gene bodies ($r = 0.23$, $p<0.001$), respectively. We identified 17,710 DM regions (3,518 hypermethylated and 14,192 hypomethylated at 70dg versus 41dg), which were enriched among gene promoters. DE genes (895 upregulated and 718 downregulated, FDR<0.05) were enriched among 3,911 DM genes. Our results agree with previously reported relationships between gene methylation and expression in other species, and reveal genes with synchronized changes in methylation and expression during pig skeletal muscle development that warrant further study.

P0443: Swine

3D Genomic Mapping Reveals Transcription Regulation and Chromatin Organization of Skeletal Muscles in Swine

Jianhua Cao, Huazhong Agricultural University, Wuhan, China

The three-dimensional (3D) chromatin organization is important for gene transcription. However, the 3D chromatin organization and its effect on gene transcription still remain largely unknown in pig genome, and the dynamics of 3D genome of skeletal muscles is much less investigated. We applied long-reads ChIA-PET strategy to comprehensively map higher-order chromatin interactions and dynamics mediated by transcription factors CTCF and RNA Polymerase II (RNAPII) in porcine longissimus dorsi muscles (LDM). We found that 1,010, 142 CTCF-defined chromatins boundaries functioned as the fundamental scaffolds for genomic 3D architectures of skeletal muscles in swine, whereas RNAPII-mediated chromatin interactions primarily served as 927,236 transcription factories to reshape the chromatin organizations in porcine myogenesis activities. Our ChIA-PET data confirmed that the boundaries delimited by CTCF-binding sites play an important role in consolidating the stability of chromatin structures via the topological associated domains (TAD). The long-range interactions within CTCF topological foci were crucial in determining gene transcription related with postnatal muscle growth. Of particular, combined with RNA-Seq data and ChIP-Seq data, the muscle specific genes like *miR-1/133*, *MSTN* have novel elucidation on 3D genome aspect with ChIA-PET data. Our findings in this study thus provide new clues and potential targets to explore key elements related to muscle growth in swine and are also likely to shed light on elucidating chromatin organization and dynamics underlying the process of mammalian myogenesis.

Key Words

Pig; ChIA-PET; 3D genome; transcription factory; chromatin organization

Highlights

- ChIA-PET has great contribution to elucidate 3D chromatin organizations of skeletal muscles in swine.
- CTCF-defined chromatin TADs and RNAPII associated transcription factories co-regulated gene transcription including skeletal muscle specific genes.
- CTCF-binding sites at genome topological boundaries play an important role in skeletal muscle growth in swine.
- A good example shows how chromatin dynamics fundamentally impact on the gene transcription in muscle specific gene transcription in myogenesis.

P0444: Swine

Molecular and Functional Variation of the Gamma Delta T Cell PRR/Co-Receptor WC1 Gene Family Among Livestock

Lauren Le Page[1], Alexandria Gillespie[1], Alehegne Yirsaw[1], Jessica Buck[1], Natalie Boisvert[1], Haeree Park[1], Julie Mclaughlin[1], Payal Damani-Yokota[1], Timothy P.L. Smith[2], Cynthia L. Baldwin[1] and **Janice C. Telfer**[1], (1)Veterinary & Animal Sciences, University of Massachusetts Amherst, Amherst, MA, (2)USDA, ARS, USMARC, Clay Center, NE

WC1, a member of the group B Scavenger Receptor Cysteine Rich (SRCR) superfamily, is expressed exclusively on γδ T cells of most mammals and birds. There are 13 genes encoding for WC1 in cattle *(WC1-1* to *WC1-13*). Previously, we determined that WC1-3, but not WC1-4, expressing bovine γδ T cells respond to *Leptospira*. This was correlated with direct WC1-3 binding to *Leptospira* via its SRCR domains. Because WC1+ γδ T cells share a restriction in their γδ TCR, and WC1 has TCR co-receptor activity, we hypothesize that WC1 co-ligation with the TCR plays the determining role in activation of WC1+ γδ T cells. Swine belong to the same order as cattle, Artiodactyl, and have WC1+ γδ T cells. We have shown that WC1+ γδ T cells in cattle respond to *Leptospira* and *Mycobacteria*, two pathogens that also infect swine. Using 5'/ 3' RACE PCR and RT-PCR, we have obtained cDNA evidence for seven WC1 genes with the SRCR domain patterns of a1-[b-c-d-e-d'] or d1-[b-c-d-e-d'] and are annotating a PacBio-sequenced genome. Through bacterial pull-down assays, we have shown that multiple SRCR domains from different WC1 genes bind to vaccine strain *Leptospira spp*, and freshly grown Pasteur and Danish strains of *Mycobacterium bovis*. Classification of WC1 genes, and their role in the activation of 𝛾δ T cells, will allow these cells to be recruited in next generation vaccines to pathogens that have significant negative economic impact.

P0445: Swine

Transcriptome Profiling in Testis of Boars with High and Low Levels of Hyperactivated Motility in Sperm

Maren van Son[1], Nina H. Tremoen[2], Ann Helen Gaustad[1], Dag Inge Våge[3], Frøydis D. Myromslien[2], Teklu T. Zeremichael[2] and Eli Grindflek[1], (1)Topigs Norsvin, Hamar, Norway, (2)Department of Natural Sciences and Technology, Inland Norway University of Applied Sciences, Hamar, Norway, (3)NMBU – CIGENE, Ås, Norway

An accurate evaluation of semen quality is essential for high quality artificial insemination in pig production. Traditionally, subjective microscopic scorings have been used to discard poor ejaculates, however, objective and precise measurements of sperm motility characteristics, morphology and concentration is now available through computer-assisted semen analysis (CASA). One of the motility characteristics recognized by CASA is sperm hyperactive motility, which is required for successful fertilization. Genetic defects during spermatogenesis can cause reduced sperm motility and male infertility, however, the molecular mechanisms underlying this process is not well understood. In this study, we examined gene expression differences in testis of pigs with high and low levels of sperm hyperactive motility. 239 different ejaculates from 104 Landrace boars were analyzed at the day of collection. Consistent measurements of at least three separate ejaculates were requested to be classified as high or low level of hyperactive motility. Testis samples from high (n=4) and low (n=4) groups of hyperactive motility were selected for transcriptome sequencing. Single reads of 100 basepair were mapped to *Sscrofa* build 11.1 and analysis revealed that 674 genes were significantly differentially expressed between the two groups (FDR < 0.05 and logFC > 1). The most significant gene was DNL-type zinc finger (*DNLZ*) (FDR = 1.2e-05) and this gene was also the most up-regulated in the high-level group (logFC = 2.7). The most down-regulated gene in the high-level group was zinc finger DBF-type containing 2 (*ZDBF2*) (logFC = -1.7). Gene ontology analysis will identify enriched pathways associated with the data.

P0446: Swine

Integration of Gene Expression Profiling of Hypothalamic Arcuate Nucleus with Genome-Wide Associations to Discover Functional Variants Associated with Age at Puberty in Gilts

Hiruni R Wijesena[1], Clay A Lents[2], Brittney N Keel[2], Jean-Jack Riethoven[1], Matt L. Spangler[1], Stephen D. Kachman[1] and Daniel C. Ciobanu[1], (1)University of Nebraska - Lincoln, Lincoln, NE, (2)USDA, ARS, USMARC, Clay Center, NE

Age at puberty (AP) in gilts is a moderately heritable trait (h^2 = 0.37) and the earliest indicator of sow reproductive longevity. Therefore, quantifying the pleiotropic sources that influence both AP and reproductive longevity is important in understanding the differences in sow fertility. In this study, we integrated genome-wide associations (GWAS), whole genome sequencing and gene expression profiling of the micro-dissected hypothalamic arcuate nucleus using RNA sequencing to identify genetic variants associated with AP in the UNL resource population (n=1,644). The arcuate nucleus plays a major role in regulating the onset of puberty through controlled secretion of gonadotropins. Seventy differentially expressed genes (DEG) were identified (P_{adj} < 0.1) between early (n=11) and late (n=6) onset of puberty gilts. Three of these genes (*CDADC1, FAM111B* and *HERPUD2*) overlapped with major (top 1%) QTL regions for AP from GWAS. Genetic variants located upstream of transcription start site (<1000 bp) affecting potential cis-binding motifs were identified as possible sources of differential expression and variation in onset of puberty. For example, SNP-affected motifs for two transcription factors (*CTCF* and *SP2*) known to regulate the expression of both *CDADC1* and *FAM111B* were identified in the proximal promoter of these genes. There were 363 upstream regulators of the 70 DEG identified. Thirty-eight upstream regulators of six DEG (*CDKN1A*, *DPP4*, *FFAR2*, *LCN2*, *PGK1* and *SAMHD1*) overlapped with major QTL regions for AP. Missense SNP were identified in four regulators (*APC*, *CDCA2, IL17B* and *RAD9A*), which could be potential trans-modulators of DEG in gilts with differences in AP.

USDA is an equal opportunity provider and employer.

Keywords: gilts, age at puberty, hypothalamus, RNA, transcription motifs

P0447: Swine

A Globin Blocker to Increase Sequencing Efficiency for QuantSeq 3' mRNA-Seq in Porcine Blood

Kyu-Sang Lim[1], Qian Dong[1], Pamela Moll[2], Jana Vitkovska[2], Gregor Wiktorin[2], Stephanie Bannister[2], Dalia Daujotyte[2], Christopher K. Tuggle[1] and Jack CM Dekkers[1], (1)Iowa State University, Ames, IA, (2)Lexogen GmbH, Vienna, Austria

The blood transcriptome is a potential source of biomarkers for many biological processes. RNAseq is a powerful method to query the blood transcriptome but is expensive. In this aspect, QuantSeq 3'RNA-seq, which generates one 3' end fragment per transcript, reduces sequencing costs and data analysis time. However, high proportions of hemoglobin (HB) mRNA of erythrocytes decreases the resolution of RNA-sequencing data from whole blood. Here, we applied a novel globin blocker assay (GB) using a HB-specific oligonucleotide mix in different

concentrations. While reads for HBA and HBB accounted for 19.4 and 36.9% of total reads in non-GB samples, GB under optimal concentrations reduced HBA and HBB reads on average to 8.7 and 2.3% of total reads. The number of genes that could be reliably detected was approximately 2200 genes greater in GB samples than in non-GB samples. The impacts of GB on globin depletion were also confirmed by differential expressed gene analysis. In conclusion, the GB can increase the resolution of QuantSeq data from porcine blood samples. GB integrates seamlessly in the QuantSeq 3'RNA-seq workflow without the need for additional reaction steps.

P0448: Swine

Association Study of Candidate and Major Genes for Growth and Fatness in Crossbred Iberian Pigs

Consolación García-Contreras[1], Miriam Ayuso[2], Almudena Fernández[1], Marta Vázquez-Gómez[3], Yolanda Núñez[1], Rita Benítez[1], Javier Garrayo[4], Susana Astiz[1], Ana I. Rey[3], Antonio González-Bulnes[1], Beatriz Isabel[3] and **Cristina Ovilo**[1], (1)INIA, Madrid, Spain, (2)University of Antwerp, Antwerpen, Belgium, (3)UCM, Madrid, Spain, (4)UAB, Barcelona, Spain

Crossbred Iberian pigs are the main genotype employed for the production of high-quality Iberian meat products, although their meat quality is lower than that observed in their pure counterparts. The objective of this work was to evaluate a wide panel of markers in known major and candidate genes, for traits involved in muscle growth, fatness and meat quality. We designed a panel of 192 polymorphisms (SNPs) employing available published information and also extracting relevant SNPs from pure and Duroc-crossbred Iberian muscle RNAseq data. Selected SNPs were located in relevant biological and functional candidate genes, regulators of pathways involved in productive traits and differentially expressed genes. The markers were genotyped in 480 animals from an Iberian x Duroc commercial population, in which phenotypes were obtained for several traits including sequential measurements of body weight (BW), average daily gain (ADG) and fatness. An association study was performed for 123 SNPs successfully genotyped and showing segregation, within the statistical animal model framework. The results confirmed the effects of several known missense mutations (such as *LEPR*, *LEP*, *MC4R* or *FTO* on fatness and growth traits or *ACSL4* on fatty acid composition) and also disclosed interesting effects of new SNPs in less known genes; such as *NMNAT1*, *SOWAHB* or *NFE2L2* genes affecting BW and ADG, *EGR1* affecting intramuscular fatty acid composition or *LRP4* affecting fatness. These results contribute to a better understanding of the genetic architecture of relevant traits for Iberian pig production and provide tools for the implementation of marker-assisted selection strategies.

P0449: Swine

Geneflow between Japanese Wild Boars and Pigs in Japan

Tomomi Aratani[1], Maho Ito[1], Keiko Takeuchi[1], Hiroshi Yasue[2], Satoru Niwata[3], Hiroshi Murata[4], Yaetsu Kurosawa[5] and Masahide Nishibori[1], (1)Hiroshima University, Higashi-hiroshima, Japan, (2)Tsukuba GeneTechnology Laboratories Inc., Wakayama, Japan, (3)Kurabo Industries Ltd., Neyagawa, Japan, (4)Hiroshima Prefectural Office, Hiroshima, Japan, (5)Tokyo University of Agriculture, Tokyo, Japan

In Japan, two subspecies of wild boar are distributed, one is Japanese wild boar (*Sus scrofa leucomystax*) and another is Ryukyu wild boar (*Sus scrofa riukiuanus*). These are reported to be genetically separated from pig. Now, damage to agriculture industries caused by wild boars in Japan is about 45 million dollars/year, and it is suggested that there is a gene flow from pig to wild boar as a factor explain the increase number of wild boars. Also, unlike wild boar, pig coat color is diverse due to domestication, and due to *MC1R* and *KIT* genes. The aim of this study is to identify the gene flow in wild boars that is associated with the livestock pigs through these genes.

As model samples, the F1 hybrids obtained by crossing Berkshire or Duroc with the wild boars, and their parents were used. In addition, the suspected gene flow in the wild boars were investigated. They were analyzed by DigiTag2 assay for 96 SNPs, and the molecular phylogenetic tree was constructed to estimate their genetic relationships. PCR-RFLP method was performed for *MC1R* and *KIT* genes.

In results of molecular phylogenetic tree and PCR-RFLP, F1 hybrids were classified in the middle of pigs and wild boars, and had inherited the characteristic polymorphisms of their parents. The gene flow in wild boars was suspected to be driven from livestock pigs confirmed by these methods. So, it is considered that these methods of this study are useful for identifying the hybrid boars among the wild boars.

P0450: Swine

Genome Sequences for African Suidae Species

Lel Eory[1], Amanda Warr[1], Heather A. Finlayson[1], Simon J. Girling[2], Karim Gharbi[3], Mick Watson[1] and Alan L. Archibald[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)The Royal Zoological Society of Scotland, Edinburgh, United Kingdom, (3)Edinburgh Genomics, University of Edinburgh, Edinburgh, United Kingdom

Wild Suidae species and indigenous domestic pigs present in Africa harbour genotypes adapted to the environmental challenges including endemic pathogens. Identification of the signatures of selection associated with these adaptive genotypes could facilitate the genetic improvement of indigenous and exotic germplasm for production in Africa. Wildlife species, such as the Common warthog (*Phacochoerus africanus*) tolerate pathogens such as African Swine Fever virus (ASFV) and act as potential reservoirs of infection. These species also represent reservoirs for parasites such as the tapeworm *Taenia solium* that cause cysticercosis in humans.

We have established draft genome sequences for three African Suidae species - Common warthog (*Phacochoerus africanus*), Red River Hog (*Potamochoerus porcus*) and Bushpig (*Potamochoerus larvatus*). For the Common Warthog we generated ~60x genome coverage in long reads using Pacific Biosciences Sequel platform from a single female individual from whom we also generated 60x genome coverage in Illumina short reads and RNA-seq data from 15 tissue samples. The long read data were assembled with Falcon to yield a draft genome of ~2.44 Gbp in 2,520 contigs with a contig N50 of 3.796 Mbp and a longest contig of 19.5 Mbp. For the Red River Hog and Bushpig we used the DISCOVAR approach based on 97x and 84x genome coverage respectively in Illumina 250 bp paired end reads. The assembly statistics are: Red River Hog 2.69 Gbp genome, 834,319 contigs, contig N50 146,579 bp, longest contig 1.69 Mbp; Bushpig 2.80 Gbp genome, 1,142,033 contigs, contig N50 95,298 bp, longest contig 1.31 Mbp.

P0451: Swine

Myostatin-Deficient Meishan Transgenic Pigs Have Increased Glucose Uptake in Stretozotocin-Induced Diabetes Status

Biao Li, Hunan Agricultural University, Changsha, China, Wentao Cui, Chinese Academy of Agricultural Sciences, Beijing, China and **Jinzeng Yang**, University of Hawaii at Manoa, Honolulu, HI

The size of skeletal muscle mass plays a significant role in glucose uptake in healthy and diabetic human subjects. Previously, we have generated myostatin-deficient ($MSTN^{-/-}$) transgenic pigs via zinc finger nucleases and cloning technology. $MSTN^{-/-}$ homozygous Meishan pigs had dramatic phenotype with individual muscle mass increase by 100% over their wild-type controls ($MSTN^{+/+}$). To understand how enhanced skeletal muscle in a large animal model are beneficial to glucose update in diabetes, we employed intravenous administration of stretozotocin (STZ) to male $MSTN^{-/-}$ and wild-type pigs (100mg/Kg body weight). One month later, blood glucose and insulin concentrations and pancreas histology were examined, STZ-induced diabetes occurred in both MSTN transgenic and wild-type pigs. The STZ-treated pigs had increased levels of fasting plasma glucose and insulin levels in comparison with animals receiving sodium citrate buffer, their pancreas also had reduced beta cells and slight increases in lymphocyte. There are significant lower concentrations of fasting plasma glucose and insulin in $MSTN^{-/-}$ pigs than that of wild-type pigs after STZ administration. Detections of pAKT and Glut4 transporter proteins by Western blotting in muscle tissue indicates significant elevations of both proteins in $MSTN^{-/-}$ pigs compared with the wild-type pigs. Therefore, the results from this large animal model suggests that enhanced skeletal muscle by manipulation of myostatin function can benefit significantly glucose uptake via AKT phosphorylation even in the status of diabetes.

P0452: Swine

Characterization and Expression Analysis of Porcine *CD163* Gene in PRRSV Infected Pigs

Byeonghwi Lim[1], Amina Khatun[2], Sang-Wook Kim[1], Salik Nazki[2], Kyung-Tai Lee[3], Choi-Kyu Park[4], Sang-Myoung Lee[5], Won-Il Kim[2] and Kwan-Suk Kim[1], (1)College of Agriculture, Life & Environmental Sciences, Chungbuk National University, Cheongju, South Korea, (2)College of Veterinary Medicine, Chonbuk National University, Iksan, South Korea, (3)National Institute of Animal Science, Rural Development Administration, Cheonan, South Korea, (4)College of Veterinary Medicine, Kyungpook National University, Daegu, South Korea, (5)College of Environmental & Biosource Science, Chonbuk National University, Iksan, South Korea

Porcine reproductive and respiratory syndrome (PRRS) is a critical viral disease that affects commercial pig productivity of swine industry worldwide. The *CD163* is a crucial candidate gene for resistance to PRRS virus (PRRSV) infection. Recently, the *CD163* gene has been proved that it is an essential factor for PRRSV infection. Therefore, the aim of this study was to characterize the porcine CD163 gene with the genomic sequence in a commercial breed and to analyze association analysis with found SNPs. A total of 65 weaning pigs were used and intramuscularly challenged with the JA142 strain. Genomic sequence of the *CD163* was identified and association analysis was performed between SNPs identified within *CD163* and host responses to viral infection that include sera viremia, lung viremia, weight gain and lung expression. the CD163 gene consisted of 16 exons and 16 SNPs were located in exon regions. Three SNPs (c.2509 G>C, c.2638 G>A and c.3534 C>T) were significantly associated with sera viremia and weight gain. The CD163 expression level also appear to be associated with these SNPs. These results suggest that the CD163 polymorphisms could be useful genetic markers for PRRSV susceptibility in pigs. Further study is underway with more animals to validate these SNPs effects and more immune cell-related traits from PRRSV infected pigs.

P0453: Swine

Effect of sncRNA on Gene Expression on the Homeostatic Status of Pigs Infected with Highly Pathogenic PRRSV

Damarius S. Fleming, ORAU, USDA National Animal Disease Center, Texas A&M, Ames, IA and Laura C. Miller, NADC-ARS-USDA, Ames, IA

It has been established that reduced susceptibility to porcine reproductive and respiratory syndrome virus (PRRSV) has a genomic component. This component, however, is a multi-faceted composition of coding and non-coding genetic elements that function as regulators of immune function. Our study focuses on the small non-coding (sncRNA) side of this response in pigs because of emergence of various sncRNAs shown to play important roles in the human viral immunity. Among these sncRNAs are the microRNA (miRNA) and transfer RNA (tRNA) molecules. Our study looks at changes in expression of these sncRNAs to produce information on how gene function in the pig can become dysregulated and subsequently respond to the virus.

The objective of the study is to identify differences in miRNA and tRNA gene expression between healthy and highly pathogenic PRRSV challenged pigs.

The study was conducted using total RNA extracted from pig whole blood taken from a total of 24 pigs split into either control (sham inoculation) or infected pigs at 1, 3, and 8 days post infection. Sequencing of the samples produced 100bp single end libraries for transcriptomic analysis of sncRNA gene expression.

The results indicated statistically significant changes in sncRNA expression were dependent on time and treatment, in which, miRNA expression was variable while tRNA expression declined steadily post-infection. The results of this study highlights changes in sncRNA expression that have the potential to unlock new targets for understanding the effect of PRRSV on pig homeostasis.

P0454: Swine

High Expression of Porcine β-Defense in *Bacillus subtilis*

Xinghua Gou, Chendu University, Chengdu, China

Porcine β-defensin (beta-defensin) is secreted by pig ,it's a small molecule antibacterial peptide which can sterilize microbe and regulate immunity. Also it's indispensable in pig's growth and reproduction .In this experiment,expression vector pHT43 was selected for expression Porcine β-defensin cDNA. When orcine β-defensin cDNA sequence was synthesized based on NCBI database, the regulatory sequences and restriction sites were added to the synthesized sequence. After it was digested by restriction enzyme, the production of digestion was inserted into expression vector pHT43for ligation. Then the production of ligation was transformated into *E.coli* DH5α for screening of the positive expression plasmid. The pHT43-SS-BD was selected by colony PCR ,and gene sequencing. The genetic engineering strain was mad by transformation of *Bacillus subtilis* WB800N with pHT43-SS-BD . When the genetic engineering strain was induced by sucrose, the molecular size of the production of iducing was identificated by SDS-PAGE, and the antibacterial activity from the inducing production was determinated

by inhibition test. SDS-PAGE electrophoresis results indicated the interest protein is about 5KDa. Inhibition test showed that the supernatant from the induced genetic engineering strain have antibacterial effect on both gram positive bacteria and gram negative bacteria. The experiment paved the way for overcoming source restrictions of antibacterial peptide,and for application of *Bacillus subtilis.*
Key words: porcine β-defensin□gene expression□antibacterial activity ;Bacillus subtilis

P0455: Cattle
Bovinemine: Simplifying Meta-Analysis in Bovine Genomic Research
Deepak R. Unni[1], Aditi Tayal[1], Darren Hagen[2] and Christine G. Elsik[1], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Oklahoma State University, Stillwater, OK
BovineMine (http://bovinegenome.org/bovinemine) is the data mining resource of the Bovine Genome Database (BGD). BovineMine enables researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. BovineMine uses the InterMine platform to integrate data from a variety of sources and makes the data accessible with simple and sophisticated data mining tools. The List Analysis and Genomic Regions search tools execute queries based on uploaded lists of identifiers and genome coordinates, respectively. These tools, combined with built-in template queries and an identifier resolver system, provide robust support for meta-analysis. For example, users can easily compare their GWAS results with published results by uploading coordinates or SNP IDs and then using the list intersection tool. Saved lists can be used in template queries for further data exploration, for example to identify published QTL overlapping a list of SNPs.
BovineMine currently includes reference genome assemblies (bovine, sheep, goat), genes (NCBI, Ensembl), proteins (UniProt), protein families and domains (InterPro), orthologs and paralogs (EnsemblCompara, Homologene, OrthoDB, TreeFam), pathways (BioCyc, KEGG, Reactome), interactions (BioGRID, IntAct), Gene Ontology (GO) annotations, QTL (AnimalQTLdb), variation (dbSNP, dbVar), SNP aliases (SNPChiMp) and publications (PubMed). Pre-computed data from BGD, including variant effects and RNAseq-based gene expression, allow users to query tissue specific gene expression levels together with genomic variation data. Future plans include the incorporation of FAANG datasets to enable fine-grained data mining of functional elements in combination with gene annotations and additional biological data.

P0456: Cattle
Non-Mendelian Genetic and Epigenetic Effects in Bovine Fetal Development
Ruijie Liu, Davies Research Centre, School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, Australia
Parent of origin-dependent genetic and epigenetic factors and their interactions are important determinants of prenatal development that remain largely unexplored. These are of considerable interest in animals and human as prenatal development impacts postnatal phenotype. We are using an intra-species large animal model based on purebred and reciprocal cross Bos taurus taurus (Bt) and B. t. indicus (Bi) that will allow us to dissect X and Y chromosome and parent of origin effects, including allelic imbalances and imprinting.
The phenotype of purebred and hybrid genome combinations includes significantly different weights and growth rates between genetic groups with pronounced sex effects have been verified. Transcriptome and miRNA data has been obtained from three fetal tissue types: Liver, Brain and Muscle. Each tissue has 3 males and females each of the 4 genetic groups (2 pure bred and 2 reciprocal crosses). Transcriptome data from liver and muscle shows there are clear separations between male and female as well as in each genetics group. We have been indicated 31 differential expression (DE) genes in liver including 10 genes are consistently significantly differentially expressed in sex comparison crossing each genetic group, which may likely to be the true sex-specific genes. Our results suggest these significantly sex effects DE genes from both sex chromosomes and autosomal may relate to the phenotype differences including liver weight and growth rates. Analyses of transcriptome data from muscle shows the similar pattern as liver. The brain is the only organ not overtly affected by genetic or sex effects on weight, revealed a different DE pattern. Further investigation will be focused on exploring the difference in transcription patterns of mRNA, small and long non-coding RNAs between tissues and also define differences in patterns of expression related to genotype.

P0457: Cattle
Putative Functional Variants in the NCAPG, ARRDC3, PLAG1 and ERGIC1 Genes Are Remarkably Associated with Pre- and Post-Natal Growth in Beef Cattle
Mahdi Saatchi, Iowa State University - American Simmental Association, Ames, IA
In beef cattle, body weight is an important trait and optimizing size at various ages is of key economic interest. In our previous study, we identified four large-effect pleiotropic QTLs located on BTA6, 7, 14 and 20 associated with body weights in several breeds. Some 96 new functional variants of candidate genes within these QTL regions were selected from the dbSNP and assayed into the new versions of GeneSeek Genomic Profilers (GGP-LD, GGP-UHD and GGP-50K). Some 5,964 Simmental, 2,982 Gelbvieh and 2,871 Red Angus animals were genotyped with one of these assays and recorded for birth, weaning and yearling weights. We performed a genome wide association study using BayesB in GenSel software. Only 9,439 in common autosomal markers were used. Variants with posterior probability of inclusion greater than 0.95 claimed as the significant markers. Some functional variants including rs109570900 encoding p.Ile442Met in NCAPG, rs109901274 encoding p.Tyr182Cys in ARRDC3, rs136369910 encoding g.25019900A>G in PLAG1 and rs43350563 encoding c.322G>A in ERGIC1 were among significantly associated markers with remarkable impact on body weights (e.g. animals with GG compare to TT genotypes for rs109570900 on BTA6 at 38777311 bp were 7 and 6 lb heavier at birth in Gelbvieh and Red Angus, 24 lb heavier at weaning in Simmental and 42 and 55 lb heavier at yearling in Simmental and Red Angus, respectively. Knowledge of such functional variants would create new opportunities for the selection of animals with appropriate body weights for harvest or maternal purposes and could decrease dystocia in beef cattle population.

P0458: Cattle
Prolactin Affects Circular RNA Profile in a Cell Line of Cattle Mammary Gland
DeWei Li, YanHuan Wang, YanHong Wang, XingTang Fang, Hong Chen and **ChunLei Zhang**, School of Life Science, Jiangsu Normal University, Xuzhou, China

Circular RNA (circRNA) is a novel type of RNA molecule generated by backsplicing, which is abundant, stable, and ubiquitously expressed. Many circRNAs have been identified with special regulatory roles. In the present study, RNAs from cow mammary gland cell line (MAC-T) were digested by RNase R, and then subjected to high throughput sequencing to investigate the circRNAs expression profile changes induced by prolactin, so as to identify the key circRNAs involved in lactation. A total of 15121 circRNAs were identified in MAC-T cells, of which 2658 circRNAs were switch on and 3402 circRNAs were switch off by prolactin. The enrichment of some Gene Ontology terms for the circRNA parental genes that induced by prolactin were also analyzed. The top 10 enriched Gene Ontology terms include reproduction, single strand break repair, regulation of DNA recombination, mitotic sister chromatid segregation, cell cycle checkpoint, DNA damage checkpoint, regulation of cyclin-dependent protein serine/threonine kinase activity, G1/S transition of mitotic cell cycle, negative regulation of transcription from RNA polymerase II promoter, establishment of mitotic spindle orientation. We further validated the prolactin-induced circRNAs cWHSC1, cDIAPH3, cCADPS2, and the downregulated cHIPK3, cKIAA0922, as well as unaffected cH2AFY, cSAMD4A by reverse transcription polymerase chain reaction and backspicing sequencing. The results indicated that circRNA might involve in prolactin induced galactosis.

P0459: Cattle

Integrated Analysis of Long Non-Coding RNAs (lncRNAs) and mRNA Expression Profiles Reveals the Potential Role of lncRNAs during Lactation Stages of Bovine Mammary Glands

Xianrui Zheng, Chao Ning, Pengju Zhao, Li Jiang and **Jian-Feng Liu**, China Agricultural University, Beijing, China

Long non-coding RNAs (lncRNAs) associated with mammary development and a number of cancers have been identified in mammals, but they are still formidable to be comprehensively identified and characterized. Here, we used RNA sequencing to profile the mammary transcriptome□lncRNA and mRNA□at peak and late lactation stages of Chinese Holstein cows. In total, more than 1000 putative lncRNAs were identified in bovine mammary, out of which 117 lncRNAs were differentially expressed between peak and late lactation stages. We observed similar genomic features between bovine mammary lncRNAs and other mammalian counterparts: with shorter transcripts, with fewer exon numbers and expressed at significantly lower levels than are protein-coding genes. Moreover, we constructed a functional lncRNA-mRNA pairs to infer function of lncRNAs. The co-expressed mRNAs were found to be associated with a wide range of biological functions, including PPAR signaling pathway, AMPK signaling pathway and mTOR signaling pathway. Further bioinformatics and integrative analyses revealed five novel lncRNAs TCONS_00303520, TCONS_00576777, TCONS_00579382, TCONS_00474808 and TCONS_00344467 played regulatory role in bovine lactation process. Our findings will facilitate the understanding of lactation process and improve the genomic resources available for cattle as well as provide a valuable resource for future bovine transcriptome studies.

Acknowledgments: We appreciate the financial support from the National Major Development Program of Transgenic Breeding (2014ZX0800953B).

P0460: Cattle

Tissue Specific ChIP-Seq Analysis of Four Histone Modifications and an Insulator Element in Bovine Adult Male Tissue Samples

Susan Margaret Waters[1], Pablo J. Ross[2] and Huaijun Zhou[2], (1)University of California, Davis, Davis, CA, (2)Animal Science, University of California, Davis, CA

The chromatin state of DNA varies from tissue to tissue, largely due to post-translational histone modifications. Histone modifications play an important role in gene expression variation between tissues and specific marks can be associated with different genome regulatory elements. H3K4me3 correlates with promoters of active genes and transcription start sites (TSS). H3K4me1 is associated with regulatory elements such as enhancers. A repressive mark, H3K27me3, identifies with promoters of inactive genes. H3K27ac is an active mark, frequently associated with active enhancers and promoters. CTCF marks the boundary between heterochromatin and euchromatin and regulates 3D chromatin structure by forming loops. Chromatin immunoprecipitation (ChIP)-sequencing is being used to annotate chromatin states for these 4 histone modifications and CTCF across multiple tissues. Eight ‘core’ tissues associated with immunity, behavior, and beef production were collected from two L1-line Hereford bulls. Tissues were snap frozen and stored at -80°C. For ChIP, 20 milligrams of each tissue is sonicated in Diagenode’s sonication buffer using the Covaris E220. DNA associated with each histone mark is collected via immunoprecipitation using Diagenode’s iDeal ChIP-Seq kit for histones. Input and sample libraries are prepared from isolated DNA using the NEBNext ChIP-Seq Library for Illumina kit from New England Biolabs, then sequenced in a HiSeq-4000 system. Annotation of these 5 chromatin marks using ChIP-seq across multiple bovine tissues will provide the information needed to identify important regulatory elements throughout the bovine genome.

P0461: Cattle

A Gene Interaction Network during Spermatogenesis in Ruminants

Wansheng Liu[1], Ti-Cheng Chang[1,2], Yuanyuan Zhang[1,3], Zhihong Liu[1,4] and Xuemei Deng[3], (1)Department of Animal Science, Penn State University, University Park, PA, (2)The Computational Biology Department, St. Jude Children's Research Hospital, Memphis, TN, (3)China Agricultural University, Beijing, China, (4)Inner Monglia Agricultural University, Hohhot, China

Spermatogenesis is a complex biological process that requires precise and highly ordered regulation of gene expression at both the transcriptional and post-transcriptional levels. The molecular mechanism underlying spermatogenesis is not completely understood. The objective of this study is to build a gene interaction network and to identify the essential functional modules in the network during spermatogenesis in ruminants. RNA-seq of the testis tissue at the age of 20 days, 8 months, and 2 years in bovine, and at the age of 15 days, 4 months, and 15 months in goats, respectively, was performed and gene annotation was conducted. A total of 2583 differentially expressed genes (DEGs) in bovine and 1538 DEGs in goats were found among the three developmental stages of the testes. Gene ontology (GO) enrichment analysis indicated that several GO terms, such as integral component of membrane, spermatid development, spermatogenesis, acrosomal vesicle, and sperm motility, were over-represented among all DEGs. A weighted gene co-expression network analysis (WGCNA) was applied to generate the co-expression networks of DEGs for the bovine and caprine testis, respectively. A parallel approach was used to build a query-query protein-protein interaction (QQPPI) network in spermatogenesis through literature mining for known proteins in the public

domains. A total of 1651 proteins related to spermatogenesis with interactions were retrieved from the IntAct database. Then, the spermatogenesis-related QQPPI network was generated and integrated with the bovine and caprine DEG co-expression network. The integrated network was used to identify functional modules during spermatogenesis in ruminants.

P0462: Cattle

Role of Age-Related Shifts in Rumen Bacteria and Methanogens in Methane Production in Cattle

Shangquan Gan, Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, China

Rumen microbiota are essential for maintaining digestive and metabolic functions, producing methane as a byproduct. Dairy heifers produce large amounts of methane based on fermentation of digested organic matter, with adverse consequences for feed efficiency and the environment. It is therefore important to understand the influence of host age on the relationship between microbiota and methane production. This study explored the age effect on the relationship between microbial communities and enteric methane production in dairy cows and heifers. Methane production and volatile fatty acid concentrations were age-related. Heifers (9-10 months) had lower methane production but higher methane production per dry matter intake (DMI) compared with adult cows. The acetate:propionate ratio decreased significantly with increasing age. Age-related microbiota changes in the rumen were reflected by a significant shift in bacterial taxa, but relatively stable archaeal taxa. *Prevotella*, *Ruminococcus*, *Flavonifractor*, *Succinivibrio*, and *Methanobrevibacter* were affected by age. This study revealed different associations between predominant bacterial phylotypes and *Methanobrevibacter* with increasing age. *Prevotella* was strongly correlated with *Methanobrevibacter* in heifers, but this association was replaced by a correlation between *Succinivibrio* and *Methanobrevibacter* in older cows (96-120 months). This shift may account for the age-related difference in rumen fermentation and methane production per DMI.

Keywords: enteric methane production, rumen microbiota, dairy cow, age-related microbiota, high-throughput sequencing

P0463: Cattle

Characterization of the Noncoding Bovine and Reindeer Rumen Papillae Transcriptome

Daniel Fischer, Ilma Tapio, Seppo Ahvenjärvi and **Johanna Vilkki**, Natural Resources Institute Finland, Jokioinen, Finland

Until now, transcriptome studies of the papillae of the rumen wall of ruminants like cattle or reindeer are rare. We present here a comparative analysis of the expression profiles of the rumen papillae based on six reindeer and eight cow samples. The transcriptomes of all samples have been sequenced with the Illumina HiSeq 3000 platform using stranded protocol with 2 x 150 bp reads. The reads were mapped against the Bovine (UMD_3.1) respective a recently published reindeer (http://gigadb.org/dataset/100370) genome. On the average, almost half of the mapped bovine reads mapped to unannotated regions of the bovine genome, indicating the need for better annotation of the genome. We analyzed long-non-coding RNA in the papillae by applying a pipeline for lncRNA-calling derived from a FAANG-Europe working group and characterize the data with respect to that. The expression profiles for non-coding transcripts are compared between reindeer and bovine. Until the publication of the reindeer genome, usually the bovine genome and annotation were used to analyze both species. Hence, with the publication of the reindeer genome and annotation, a special focus of this study is on the effect of the new, species-specific genome and its effect on the results of the characterization. For that, we analyze for the reindeer data also the differences between using the bovine genome as a reference versus using the new reindeer specific genome.

P0464: Cattle

Potential Role of Maternal Nutrition during Late Gestation on Early Calf Rumen Microbiome

Kathleen J. Austin[1], Hannah C. Cunningham[1], Kristi Cammack[2], John C. McEwan[3], Christina D. Moon[4] and Alan McCulloch[3], (1)Department of Animal Science, University of Wyoming, Laramie, WY, (2)South Dakota State University, Brookings, SD, (3)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (4)AgResearch, Grasslands Research Centre, Palmerston North, New Zealand

The maternal fetal interface may be a platform to optimize production efficiency and sustainability because of the influence of maternal factors on fetal development and offspring performance. Additionally, maternal plane of nutrition affects fetal development as well as performance later in life. Our objective was to determine the influence of maternal diet during late gestation on the early calf rumen microbiome. We hypothesized that perturbations to offspring performance resulting from maternal nutrient restriction, may be in part, due to alterations in the rumen microbiome. Cows were fed the same diet to meet or exceed NRC requirements (CON; n = 6) or to reduce body condition by 1.5 - 2 points using feed restriction (NR; n = 6) during the last 3 months of gestation. Calves were reared with their dam until weaning. Shotgun metagenomic sequencing was performed on DNA from rumen fluid collected from calves born to CON cows (n = 4) and NR cows (n = 6) at 7 days of age. The 16S rRNA gene sequences were extracted utilizing Bowtie2 against the SILVA database, and the distribution of these in each sample was analyzed using QIIME. Results indicated no differences in alpha or beta diversity, however the abundance of 8 genus-level taxa differed ($P < 0.05$) between the two groups. These results suggest that maternal plane of nutrition during gestation does not appear to have a large influence on calf rumen microbiome diversity.

P0465: Cattle

Effects of Mode of Delivery on Rumen Microbiome of Week Old Calves

Hannah C. Cunningham[1], Kathleen J. Austin[1], Kristi Cammack[2], John C. McEwan[3], Christina D. Moon[4] and Alan McCulloch[3], (1)Department of Animal Science, University of Wyoming, Laramie, WY, (2)South Dakota State University, Brookings, SD, (3)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (4)AgResearch, Grasslands Research Centre, Palmerston North, New Zealand

While it is widely accepted that the rumen microbiome is sterile at birth, there is evidence that colonization may actually be initiated at or prior to birth. Furthermore, there is evidence that mode of delivery influences infant health due to changes in the human infant gastrointestinal microbiome. Colonization of the gastrointestinal tract by beneficial microbes is important for regulation of the immune system, and ultimately calf health. Initial colonization of the rumen is a critical stage that affects the establishment of fermentation in the pre-ruminant; microbial production of volatile fatty acids subsequently enhances rumen development. We hypothesized that mode of delivery would result in

differences in the rumen microbiome of young calves. Our objective was to determine if calves born via caesarean harbored a different rumen microbiome compared to calves born vaginally. Shotgun metagenomic sequencing was performed on DNA from rumen fluid collected from caesarean (n = 4) and vaginally (n = 6) born calves at 7 days of age. The 16S rRNA gene sequences were extracted utilizing Bowtie2 against the SILVA database, and the distribution of these in each sample was analyzed using QIIME. Results indicate increased species richness ($P < 0.05$) in vaginally delivered calves compared to caesarean delivered calves. Additionally, significant differences in the abundances of 16 genus-level taxa were observed between the two groups. Results suggest that mode of delivery does affect the rumen microbiome in pre-ruminant calves. This may have implications in herd health and production that could persist through maturity.

P0466: Cattle

Integrative Analysis of Transcriptome, Epigenome and GWAS Data Enhances the Understanding of the Genetic Basis Underlying Rumen Development in Dairy Cattle

Lingzhao Fang, University of Maryland at College Park, College Park, MD, Li Ma, Department of Animal and Avian Sciences, University of Maryland, College Park, MD, George E. Liu, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD and Cong-jun Li, ARS, USDA, Beltsville, MD

A well-developed, fully functional rumen ensures a dairy calf to become an efficient and productive animal. A better understanding of the genetic and biological mechanisms underlying the rumen development, especially before and after weaning, is important, yet it is poorly elucidated. In this study, rumen samples from four Holstein calves were collected before and after weaning respectively. The transcriptome, three histone marks (H3K9, H3K27, and H3K9me3,), and RNA pol II were genome-wide investigated using RNA-Seq and ChIP-Seq technologies, respectively. Differentially expressed genes and modified histone marks were detected under the comparison of before *vs.* after weaning. Integrative analysis of gene expression and histone mark modification was performed to detect genomic features that were involved in the rumen development. Sequence-based genome-wide association studies for 38 complex traits of economic importance were further integrated to assess the enrichment of association signals with the detected genomic features. This study provided great insights into the genetic mechanisms underlying the rumen development before and after weaning, and offered a general framework to integrating multi-layers data to investigate the biological and genetic basis underpinning complex traits.

P0467: Cattle

FANA_Antisense Oligonucleotides: Next Generation RNA Silencing Technology to Target Pathogens in Blood and Ectoparasites

Wayne Hunter[1], Jackie Lin Metz[2,3], Kevin B. Temeyer[4], Beto Perez de Leon[4], John Tomich[5] and Veenu Aishwarya[6], (1)USDA Agricultural Research Service, Fort Pierce, FL, (2)University of Florida, Ft. Pierce, FL, (3)AUM LifeTech, Inc., Philadelphia, PA, (4)USDA, Kerrville, TX, (5)Kansas State University, Manhattan, KS, (6)AUM LifeTech, Inc, Philadelphia, PA

Solutions are needed for an extensive set of clinical and agricultural animal problems ranging from eliminating ectoparasites of livestock to curing patients of HIV and multidrug-resistant bacteria. However, stability to nuclease degradation, off-target effects, cytotoxicity, and intracellular delivery of gene targeting products pose barriers to the effective translation of these technologies. Here we report FANA modified antisense oligonucleotides, 2'-deoxy-2'-fluoro-D- arabinonucleic acid antisense oligonucleotides (FANA ASOs) as a potential solution to these problems. Studies with FANA ASOs have shown them to inhibit replication of HIV in human peripheral blood mononuclear cells. FANAs work in a wide spectrum of cellular and animal models and can knockdown any desired RNA target. Successful intracellular delivery of FANA ASOs required no delivery agents (conjugates or formulations) and no cytotoxicity or off-target effects were observed. Furthermore, FANA ASOs spiked into whole mammalian blood, or blood sera, showed increased resistance to nuclease degradation beyond 6 days. FANA's also maintained high inhibitory activity to their designed target transcripts at 6 days post treatment in blood or sera, and this was increased further when combined with BAPC technology. BAPCs (Branched Amphiphilic Peptide Capsules) are a new class of self-assembling peptide nano-capsular spheres (Phoreus Biotechnology, Inc.). Peptide based nano-assemblies show promise as nano-delivery vehicles for the safe, targeted transport of drugs to specific tissues and organs with minimal off target accumulation. These results suggest significant potential for application of FANA ASOs to address issues of human and animal health as treatments to reduce pathogens, as well as arthropod vectors.

P0468: Cattle

Analysis of Small Non-Coding RNA Profiles Resulting from a *Pasteurella multocida* Challenge in Cattle

Tasia M. Taxis[1], Fred M. Tatum[2], Robert E. Briggs[2] and **Eduardo Casas**[2], (1)Michigan State University, East Lansing, MI, (2)USDA, ARS, National Animal Disease Center, Ames, IA

Small non-coding RNAs (sncRNAs), such as microRNAs and tRNA-derived fragments (tRFs) have been linked with immune response. The objective was to assess changes in microRNAs and tRFs associated with virulent *Pasteurella multocida* (a bovine respiratory pathogen) challenge in calves previously exposed to isogenic *P. multocida* strains. Three modified *P. multocida* strains were produced by removing one of three genes encoding putative virulence factors. Cattle were intranasal-inoculated with $2x10^9$ cfu/mL as follows: Control (sham inoculation; n=3), Wt (inoculated with wild-type *P. multocida*; n= 4), FhaB2 (n= 4), HyaE (n=4), and NanP (n= 4). Fifty six days later all animals were intratracheally challenged with $2.2x10^9$ cfu/mL of wild type *P. multocida*. ELISAs for immunoglobulin levels and bacterial shedding were assessed weekly and showed significant differences between exposed and sham groups. Small noncoding RNAs were sequenced from sera and white blood cells from all animals before and after intratracheal challenge (days 49 and 61, respectively). A significant interaction of treatment and challenge was detected for bta-miR-150 ($P= 0.0025$) and $tRF5^{MetCAT}$ ($P= 0.0009$). For bta-miR-150, Control and FhaB2 counts decreased after challenge, Wt and NanP increased, and no difference was observed for HyaE. For $tRF5^{MetCAT}$, Control, FhaB2, and HyaE counts increased, and Wt and NanP decreased after challenge. While Control and FhaB2 counts decreased for bta-miR-150, they increased for $tRF5^{MetCAT}$. An inverse relationship between bta-miR-150 and $tRF5^{MetCAT}$ was observed. Results indicate that variation in calf response to *P. multocida* with specific genomic modifications leads to differences in host's counts of sncRNAs.

P0469: Cattle

MicroRNA Expression Profiles of Bovine Monocyte-Derived Macrophages Infected *in vitro* with Two Strains of *Streptococcus agalactiae*

Anna Monika Lewandowska-Sabat[1], Silje Camilla Furre Hansen[1], Trygve Roger Solberg[2], Olav Østerås[3], Bjørg Heringstad[4], Preben Boysen[1] and **Ingrid Olsaker**[1], (1)Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway, (2)Geno Breeding and A.I. Association, Hamar, Norway, (3)Norwegian Cattle Health Services and TINE Extension Services, Ås, Norway, (4)Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

MicroRNAs (miRNAs) are short, non-coding RNAs, playing a key role in the regulation of innate and adaptive immune responses. In a subclinical infection such as bovine streptococcal mastitis, early recognition is a challenge, and miRNA profiling may contribute to diagnosis and to understanding of pathogenicity and defense mechanisms.

We have examined the miRNA repertoire (by sequencing of 20 libraries of small RNA) and the transcript levels (by RT-qPCR) of six key immune genes [*tumor necrosis factor alpha* (*TNFα*), *interleukin* (*IL*)*-1 beta* (*IL-1β*), *interleukin-6* (*IL-6*), *interleukin-8* (*IL-8*), *interleukin-10* (*IL-10*) and *transforming growth factor beta 1* (*TGFβ1*)] in bovine blood monocyte-derived macrophages exposed *in vitro* for 6 hours to two live sequence type strains of *Streptococcus agalactiae* (ST103 and ST12).

Comparing high quality sequence reads from unchallenged controls with exposed macrophages revealed that 44 and 17 miRNAs were differentially expressed ($P < 0.05$) after exposure to ST12 and ST103, respectively. We also identified expression of 31 potentially novel bovine miRNAs. Pathway analysis of the differentially regulated miRNAs and predicted target genes in the macrophages infected with ST12 revealed significant enrichment for inflammatory response and apoptosis, while significant enrichment for integrin and GABA signaling were found in the ST103 infected macrophages. Both bacterial strains regulated miRNAs involved in the alternative activation of macrophages. Furthermore, *TNF-α*, *IL-1β*, *IL-6*, *IL-8* and *IL-10* transcripts were significantly upregulated by both bacterial strains, however only strain ST12 caused significant downregulation of the *TGFβ1* transcript.

Hence, differential pathogen-induced regulation of miRNAs in bovine macrophages controlling inflammation were identified.

P0470: Cattle

Comparative and Integrative Genomics of Bovine and Human Tuberculosis: A One Health Perspective

David E. MacHugh, University College Dublin, Dublin, Ireland

Human tuberculosis, caused by *Mycobacterium tuberculosis*, is second only to HIV/AIDS as the greatest killer worldwide due to a single infectious agent. Bovine tuberculosis, caused by the closely related *Mycobacterium bovis* (99.95% sequence identity), is an economically-important disease affecting global cattle production, particularly in many developing countries where it also represents a significant zoonotic disease.

We have taken a One Health approach to tuberculosis by using network-based approaches to compare and integrate host transcriptome data from human and bovine macrophages infected with *M. tuberculosis* and *M. bovis*, respectively. These analyses have shed light on host-pathogen interaction for human and bovine tuberculosis disease and have provided information that can also be used to prioritize genome-wide association data and enhance detection of genomic variants for susceptibility/resistance to mycobacterial infections in cattle and humans. In addition, we have used high-throughput epigenomics to functionally dissect transcriptional control of the bovine alveolar macrophage response to infection with *M. bovis*. Results from this work reveal mechanisms underlying mammalian macrophage M1/M2 polarisation in response to mycobacterial infection and provide a novel perspective on host-pathogen interaction. Finally, we have demonstrated that peripheral blood transcriptomics can provide a route to development of novel biomarkers for mycobacterial infections in cattle with significant implications for diagnosis of human tuberculosis.

P0471: Cattle

A Nutrigenomic Perspective to Search for Gene Variants That Influence Carcass Traits of Feedlot Cattle

Kelsey E. Hamilton, Kayla M Madder and Fiona C Buchanan, University of Saskatchewan, Saskatoon, SK, Canada

Vitamin A has a nutrigenomic effect on intramuscular fat. Discovering variants in genes involved in fat deposition and affected by vitamin A could allow feedlots to precision feed to optimise carcass traits. Genes previously shown to be affected by vitamin A (*ANPEP*, *CLU*, *ADFP, GPX3*, *SPARC*, and *IGFBP6*) were sequenced, and *ANPEPc.410G>A* was selected for genotyping in a population of mixed breed steers (N=988). This population was fed vitamin A at 100% or 50% NRC recommended levels. Yield ($P<0.01$; *AA*=2.47±0.03, *GA*=2.36±0.03, *GG*=2.14±0.08), marbling score ($P<0.01$; *AA*=397.2±2.7, *GA*=388.6±3.3, *GG*=370.4±7.2), and fat ($P<0.01$; *AA*=8.52±0.17 *GA*=7.58±0.21 *GG*=7.04±0.46) were all significantly affected by the *ANPEP* variant. Vitamin A also had an effect on fat ($P<0.05$), and an interaction between *ANPEP* and vitamin A affected rib-eye area ($P<0.05$). The *ANPEP* SNP was genotyped in a second population of mixed breed steers (N=708) fed a standard feedlot ration with the NRC recommended level of vitamin A. Significance was found with yield, marbling, fat, and rib-eye area. The *AA* genotype was significantly more marbled, while *GG* animals were leaner with better yields. Interestingly, *ANPEPc.410G>A* is in exon one and is the fourth variant in a haplotype containing 11 SNPs that are in linkage disquilibrium. This was confirmed by sequencing cattle of various breeds, in different populations spanning many years. The three variations of this haplotype could affect gene expression by altering transcription or translation efficiency. Investigation of the functional effects of this haplotype needs to be completed in order to understand how it alters feedlot cattle performance.

P0472: Cattle

Analysis of Genetic Markers Previously Associated to Feed Efficiency Traits in Mexican Beef Cattle

Estela Garza Brenner[1], Ana M. Sifuentes Rincón[1], Felipe A. Rodriguez Almeida[2] and Williams Arellano Vera[1], (1)INSTITUTO POLITECNICO NACIONAL/CBG Laboratorio de Biotecnologia Animal, Cd. Reynosa, Tam, Mexico, (2)UNIVERSIDAD AUTONOMA DE CHIHUAHUA/FACULTAD DE ZOOTECNIA Y ECOLOGIA, Chihuahua, Mexico

Given the importance of feed efficiency traits for livestock production, efforts to implement their measurement in Mexican beef cattle have been made lately. To evaluate the effect of markers LEP, GH-*Alu* and IGF1/S*na*B1, previously reported to be associated to feed efficiency and production traits, young bulls (273 ± 38 d) of three breeds (Angus n=63, Brangus n=41 and Charolais n=32) in a 70-d trial of a centralized

performance test were genotyped. After 20 d of adaptation, daily feed intakes measured with a GrowSafe® system and body weight measured every 14 d, were registered. Average daily gain (ADG), feed to gain ratio (F:G), residual feed intake (RFI) and dry matter feed intake (DMFI) were determined and statistically analyzed with SAS software. Allelic frequencies for all markers were higher for the allele considered as favorable (GH-*Alu*, C:0.66-0.78; IGF1/S*na*B1: B, 0.55-0.67 and LEP: T, 0.59-0.72). Brangus, bulls showed a significant association (P = 0.05) of the IGF/SnaB1 marker to F:G, with higher values for both BB and AB genotypes (6.41±0.21 y 6.12±0.18 kg/kg, respectively) than for the homozygotes AA bulls (5.26±0.37 kg/kg).Our results are consistent with some previously reported in beef cattle breeds, where a dominance effect of the B over A allele was observed to affect some productivity traits. Further studies are needed to confirm the F:G association considering a larger number of individuals and other cattle breeds, in order to integrate genetic markers information as part of the selection criteria to complement the genetic evaluations efforts on efficiency phenotyping in Mexico.

P0473: Cattle

Prevalence of DNA Methylation in the Bovine Brain Methylome

Bonnie Cantrell[1], Nathan Jebbett[1], Robert C. Switzer III[2], Eugene Delay[1], Steve A Zinn[3], Sharon Aborn[4], Jane O'Neil[1], Rick Funston[5], Robert Weaber[6] and Stephanie McKay[1], (1)University of Vermont, Burlington, VT, (2)NeuroScience Associates, Inc., Knoxville, TN, (3)Department of Animal Science, University of Connecticut, Storrs, CT, (4)University of Connecticut, Storrs, CT, (5)University of Nebraska, North Platte, NE, (6)Kansas State University, Manhattan, KS

Phenotypic variation in complex traits is regulated by genetics and epigenetic mechanisms. Our research focuses on determining the extent to which DNA methylation may affect phenotypic variation in docility. Whole genome bisulfite sequencing (WGBS) was undertaken in five functionally distinct brain regions for each of eight Red Angus X Simmental steers with extreme measures of docility (n=4 per group). Whole genome bisulfite sequencing was used to gain a single base pair resolution characterization of the methylomes of the amygdala, cingulate gyrus, hippocampus, periaqueductal gray, and prefrontal cortex. An average of 32 million sequence read pairs were generated from each sample. Reads were quality trimmed with trim galore using a phred score of 20. Trimmed reads were aligned to a bovine reference index using BS seeker2 and the bowtie2 aligner. BS seeker2 was then used to measure DNA methylation levels for methylated and non-methylated CGs, as well as methylated CHG (H= A, C or T) and CHH. Next, global percentages of DNA methylation was measured for CG, CHG, and CHH by tissue type. Further analysis was undertaken to determine the prevalence of each base in the in the H context for CHG and CHH. To the best of our knowledge, this project is the largest WGBS data set for brain tissues in the world and it only begins to explain the relationship between DNA methylation, the bovine brain, and phenotypic variation for extreme measures of docility.

P0474: Cattle

Investigation of Differentially Expressed Transcripts in Cattle Supplemented with β-Adrenergic Agonists

Rachel M. Kubik, University of Nebraska - Lincoln, Lincoln, NE, Dustin T. Yates, University of Nebraska - Lincoln, Ty B. Schmidt, University of Nebraska- Lincoln, Lincoln, NE and Jessica L. Petersen, University of Nebraska, Lincoln, NE

β-adrenergic agonists (β-AA), commonly fed to cattle during the last 20-40 days of the finishing period, improve muscle growth by decreasing adipose deposition and increasing muscle accretion. β-adrenergic agonists act through specific 7-transmembrane β-adrenoreceptors and are classified by the receptor isoform to which β-AA's primarily bind. Two β-AA, Ractopamine HCl (β1-AA) and Zilpaterol HCl (β2-AA) are currently approved for use in beef cattle in the United States. The purpose of this study was to quantify differential gene expression after supplementation with either a β1-AA or β2-AA. Based on differential gene expression data, pathway analysis was performed to identify specific pathways altered by supplementation. To examine the impact of β1-AA on the blood transcriptome, whole blood samples were drawn from eight steers and eight heifers prior to and post a 20-day supplementation period which included Ractopamine HCl. The impact of β2-AA on the muscle transcriptome was evaluated from biceps femoris and latissimus dorsi sampled at harvest from 11 heifers, six supplemented with Zilpaterol HCl and five, non-supplemented controls. RNA was isolated from all samples, sequenced through 3' Tag-Seq, and after quality control, analyzed for differentially expressed transcripts. This work will help provide an understanding of the pathways that are affected by β-AA supplementation, which could lead to the development of new feed supplements or management practices to maximize production efficiency.

P0475: Cattle

Hair Coat and Thermoregulation in Brangus Heifers

Heather M. Hamblen, Adriana Zolini, Peter J. Hansen, Toni Oltenacu and Raluca Mateescu, Department of Animal Sciences, University of Florida, Gainesville, FL

Heat stress is a major cause of economic loss for beef cattle producers in tropical and subtropical environments, as the animal's true genetic potential may not be expressed. Integrating *Bos Indicus* genetics into herds has improved the heat adaptability in crossbred animals but it has also introduced other challenges related to meat quality and reproduction. The overall goal of this study is to develop genomic tools to be used in selection programs to improve heat tolerance, while also improving production traits. Hair type is an important factor influencing thermoregulation in cattle, as it insulates the body, making heat exchange less efficient. The variation of length and thickness of hair both between and within breeds suggests the selection for a coat advantageous for improved thermotolerance in cattle is possible. Coat color, coat scores and daily body temperatures at 5-min intervals over a 5-day period were recorded on approximately 725 Brangus two-year old heifers during the summer of 2016. The coat was scored as excessively smooth (score 1, n = 526), fairly smooth (score 2, n = 189) or long coat (score 3, n= 7). A repeated measures mixed model was used to investigate the effect of coat score on body temperature and was shown to be significant (P <.0001). Heifers with an excessively smooth coat had lower body temperatures throughout the 3 full days of continuous body temperature measurements, indicating that coat type plays an important role in the control of body temperature.

P0476: Cattle

Genome-Wide Association and Gene Enrichment Analyses of Meat Sensory Traits in a Crossbred Brahman-Angus Population

Joel David Leal, Mauricio Elzo, Dwain Johnson and Raluca Mateescu, Department of Animal Sciences, University of Florida, Gainesville, FL
Genome-wide association procedures combined with gene enrichment analyses (GEA) are used to identify regions of large effect, but also small effect regions with possible additive genetic effects. The objective of this study was to apply this methodology to meat sensory traits in an Angus-Brahman cattle population. Tenderness (TN), connective tissue amount (CT), juiciness (JC), marbling (MR) and flavor (FL) were measured through a trained sensory panel on 496 steaks. Animals were genotyped with the Bovine GGP-F250 array. Data processing and analysis were performed using the Genetics Q-K workflow of JMP-Genomics 6.0 software, applying the mixed model K method. The GEA was carried out using an in-house scripts written in JAVA. All SNPs with uncorrected p-value≤0.05 were included in the GEA. Gene enrichment was performed using the hypergeometric test. A total of 95, 86, 89, 81, and 92 pathways for FL, CT, MR, JC, and TN, respectively, were included in the analysis. Forty seven genes were determine as associated (p-value<0.0001) with at least one trait, and 30% of these genes were related with gene expression, 17% with cell-signaling, and 11% with cell differentiation. One polymorphism in ANO2 gene and another one in MMRN2 gene reached the genome wide adjusted significance (p-value≤6e-07) for CT and JC, respectively. In the GEA, three pathways were identified as enriched: the "Negative regulation of transcription from RNA polymerase II promoter" pathway in CT, "Endoplasmic reticulum membrane" in TN, JC, and CT, and "Positive regulation of transcription from RNA polymerase II promoter" in FL and CT.

P0477: Cattle

A Genome-Wide Association Study Identified SNCA as a Candidate Gene Related to Temperament in Angus and Brangus Cattle

Francisco A. Paredes-Sánchez[1], Ana M. Sifuentes Rincón[1], Estela Garza Brenner[1], Eduardo Casas[2], P. Ambriz-Morales[1], Felipe A. Rodriguez Almeida[3], G. M. Parra-Bracamonte[1] and R. D. Randel[4], (1)INSTITUTO POLITECNICO NACIONAL/CBG Laboratorio de Biotecnologia Animal, Cd. Reynosa, Tam, Mexico, (2)USDA, ARS, National Animal Disease Center, Ames, IA, (3)UNIVERSIDAD AUTONOMA DE CHIHUAHUA/FACULTAD DE ZOOTECNIA Y ECOLOGIA, Chihuahua, Mexico, (4)Texas Agricultural Experiment Station, Overton, TX
Cattle temperament, defined as the animal's response to handling, has been studied in different breeds, and has been shown to affect growth, health, and carcass merit. Molecular studies aimed at defining temperament in cattle are scarce. The objective was identify genomic regions associated with temperament in beef cattle. Angus and Brangus (n=104; 273±38 d) bulls were included in the study under a centralized feed efficiency performance tests based on residual feed intake (RFI) in northern Mexico. Temperament was assessed using exit velocity (EV), measuring the rate of travel over a distance of 1.83-m with an infrared sensor [EV=distance (m)/time (s)]. Based on EV, selective genotyping was achieved selecting a group of the most calm (n=16) and the most temperamental animals (n=17). Animals were genotyped using the GeneSeek Genomic Profiler HD 150K chip. After quality control 139,376 SNPs were used for a case-control association analysis. Markers rs133956611 (p-value= 2.65 E-06) and rs81144933 (p-value 9.58 E-06), were associated with cattle temperament. The rs133956611 maps to over 222 kb upstream to synuclein alpha (SNCA) and multimerin 1 (MMRN1) genes on BTA6; rs81144933 maps to over 344 kb downstream to GPRIN family member 3 (GPRIN3) on the same chromosome. SNCA seems to be a strong candidate to predict temperament. The literature suggests that SNCA plays an important role in maintaining an adequate supply of synaptic vesicles in presynaptic terminals. It may also help regulate the release of dopamine, that is critical for controlling the start and stop of voluntary movements.

P0478: Cattle

Alternative Transcript Splicing Events in *Longissimus dorsi* Muscle of Nelore Cattle Phenotypically Divergent for Ribeye Muscle Area

Danielly B. S. Silva[1], Larissa Fernanda S. Fonseca[1], Daniel G. Pinheiro[1], Maria Malane M. Muniz[1], Ana Fabrícia B. Magalhães[1], Lúcio Flávio Macedo Mota[2], Gerardo A. Fernandes Júnior[2], Luis Artur L. Chardulo[3], Robert D. Schnabel[4], Jeremy F. Taylor[5] and Lucia Galvão de Albuquerque[6], (1)University of São Paulo State/UNESP, Jaboticabal, Brazil, (2)School of Agricultural and Veterinarian Sciences - São Paulo State University, Jaboticabal, Brazil, (3)University of São Paulo State/UNESP, Botucatu, Brazil, (4)Division of Animal Sciences, Informatics Institute, Columbia, MO, (5)Division of Animal Sciences, University of Missouri, Columbia, MO, (6)School of Agricultural and Veterinarian Sciences - São Paulo State University and National Council for Science and Technological Development, Jaboticabal, Brazil
Alternative pre-mRNA splicing is known to affect the majority of bovine genes that play important roles in development and muscle growth. Our aim was to use RNA-Seq to identify differentially expressed alternatively spliced (DAS) transcripts in Nelore *Longissimus dorsi* muscle. From a total of 80 animals phenotyped for rib eye area (REA), 15 with the highest REA (HREA) and 15 with the lowest REA (LREA) were used. The JuncBASE package was used to identify alternatively spliced transcripts and a total of 13,986 alternative splicing events were identified between the HREA and LREA groups. Of these, 258 transcript isoforms were DAS ($p \leq 0.05$), with 31.8% (82) representing exon skipping; 30.6% (79) alternative 3' splice sites; 24% (62) alternative 5' splice sites; 5.8% (15) alternative first exons; 2.7% (7) coordinate exon skipping; 2.7% (7) intron retention; 1.6% (4) alternative last exons and 0.8% (2) mutually exclusive exons. These DAS were transcribed from 206 genes with roles in metabolic pathways related to muscle growth and development, such as, SHC-related events triggered by IGF1R, signals to RAS and RAF/MAPK cascades. We identified three key-player genes underlying the DAS: *DUSP13*, *RPS27L* and *DNM2*. These genes play roles in cell regulation and differentiation with *DNM2* being involved in the production of microtube bundles as well as binding and hydrolyzing GTP. The multiple different types of alternative splicing events and large number of genes associated, producing DAS transcripts, show the complexity of the muscle development transcriptome in Nelore cattle.
Acknowledgments: São Paulo Research Foundation FAPESP (grants #2009/16118-5 #2015/16850-9 and #2016/22894-1)

P0479: Cattle

Genotype x Environment Interaction Analysis Reveals Genomic Regions Associated with Nelore Cattle Scrotal Circumference

Lúcio Flávio Macedo Mota[1], Tiago Passafaro[2], Gerardo A. Fernandes Júnior[1], Rafael Espigolan[1], Tiago Bresolin[1], Guilherme J. M. Rosa[2], Roberto Carvalheiro[3] and Lucia Galvão de Albuquerque[3], (1)School of Agricultural and Veterinarian Sciences - São Paulo State University, Jaboticabal, Brazil, (2)University of Wisconsin-Madison, Madison, WI, (3)School of Agricultural and Veterinarian Sciences - São Paulo State University and National Council for Science and Technological Development, Jaboticabal, Brazil

Knowledge of genomic regions involved with genotype by environmental interactions in livestock is extremely important to aid the development of breeding programs in harsh conditions and with climate change. This study aimed to identify genomic regions and causal networks associated with scrotal circumference (SC, cm) at yearling in Nelore cattle, raised under different environmental conditions (EC). The SC of 151,553 animals and genotypic information on 416,555 SNP markers of 3,126 young bulls and 810 sires were used. Animals were raised on farms located in three different Brazilian regions (Midwest, Southeast and Northeast). A two-step reaction norm model was applied to estimate the response of animals to the environmental changes. The SNP effects were estimated for SC in three EC levels (Low, Intermediate, and High) and the causal networks were inferred for significant SNP markers using a Bayesian network approach applying the Incremental Association Markov Blanket algorithm. SNP effects were deemed significant using $-\log_{10}(\text{q-value})>5.0$, where q-value is an adjusted p-value based on false discovery rate. Genes located within ±100 kb of each SNP associated with SC across EC were identified using Ensembl. Bayesian network analyses identified potential causal effects of SNPs associated with variations in SC, as well as associations between SC across EC. A total of eight SNPs located on chromosomes 8, 19, 23 and 25 were associated with SC in different EC. Candidate genes in these regions play an important role in the regulation of biological process associated with decreased testis weight, and degeneration and atrophy of the seminiferous epithelium. These results contribute to a better understanding of the molecular mechanisms affecting SC in Nellore cattle raised under different environmental conditions. Acknowledgments: São Paulo Research Foundation (FAPESP) grants #2009/16118-5 and #2017/02291-3

P0480: Cattle

miRNA and mRNA Co-Expression Networks Associated to Feed Efficiency in Nelore Cattle

Priscila S.N. Oliveira[1], Polyana C. Tizioto[2], Aline S.M. Cesar[2], Gabriella Borba de Oliveira[2], Wellison Jarles Silva Diniz[1], Luiz L. Coutinho[2], James M. Reecy[3] and Luciana C.A. Regitano[4], (1)Embrapa Pecuária Sudeste, São Carlos, Brazil, (2)University of São Paulo, Piracicaba, Brazil, (3)Department of Animal Science, Iowa State University, Ames, IA, (4)Embrapa Pecuaria Sudeste, São Carlos, Brazil

Residual Feed Intake (RFI) is a measure of feed efficiency. In addition to consuming less feed, efficient animals produce fewer pollutants, e.g. methane. Here, we performed gene co-expression network analysis to uncover the complex interactions between miRNAs and mRNAs that are expressed in skeletal muscle and liver tissue of Nelore steers (n=30). miRNA and mRNA expression data were obtained from *Bos indicus* cattle that had extreme BLUP (Best Linear Unbiased Prediction) values for RFI. After quality control and data normalization, Weighted Gene Co-expression Network Analysis (WGCNA) was completed. Two miRNA modules identified by WGCNA were correlated with RFI in the more efficient animals; ME*blue* (pvalue=1.44e-02) and ME*pink* (pvalue=1.41e-04). Genes in these modules were associated with the Insulin signaling pathway. Moreover, after correlating miRNA with mRNA modules, one of the correlated mRNA modules (ME*blue*; r = -0.5; pvalue=0.05) was enriched for genes related to Insulin resistance. Enrichment analysis of target genes of bta-miR-486; previously identified as downregulated in skeletal muscle of this same set of samples; also indicated that the Insulin signaling pathway was overrepresented. It's known that the insulin metabolism influences feed efficiency related traits in cattle. However, the miRNA - mRNA interactions identified here are new and may help elucidate important unknown regulatory mechanisms of feed efficiency in Nelore cattle. These results indicate that miRNAs and genes from these modules can regulate biological processes involved in insulin metabolism. Additionally, further studies will help to elucidate the molecular mechanisms that control feed efficiency in cattle.

P0481: Cattle

Histological Muscle Characterization in Hypertrophied Marchigiana Beef Cattle Breed

Domenico Aiello[1], Simone Ceccobelli[2], Guglielmo Sorci[3] and Emiliano Lasagna[2], (1)University of Perugia, Perugia, ID, Italy, (2)Department of Agricultural, Food and Environmental Sciences - University of Perugia., Perugia, Italy, (3)Department Experimental Medicine - University of Perugia, Perugia, Italy

Myostatin (MSTN) is one of the major regulators of skeletal muscle development. Mutations on MSTN gene are responsible for double muscling phenotype in several livestock animals. This phenotype occurs at a high frequency in some breeds like the Italian Marchigiana beef cattle breed, where it was identified a transversion mutation (g.874G>T) in exon 3 of the MSTN gene.

In this work a PCR-RFLP test was performed to determine the genotype of ten bullocks at MSTN locus. To better understand the myostatin function histological analysis were carried out to investigate on differences in muscle morphology between the three genotypes (GG,GT, GG). Three different muscles (Semitendinosus, Psoas major and Longissimus dorsi) were selected and collected at slaughtering for the characterization and the myofiber Cross Section Area (CSA) was used for the comparison. The CSA values in homozygous (TT) and normal (GG) were quite similar but microscopy analysis revealed muscle hyperplasia in homozygous (TT) bullocks, in accordance with previous studies.

Furthermore, considering the important role played by myostatin during myogenesis, a satellite cell specific marker (PAX7) was considered. Part of each muscle samples were analysed for the expression of PAX7 by Western blotting. Results showed different expression levels of PAX7 in all the three genotypes, with an increase in the mutant homozygous one. These findings confirm that lack of myostatin influences the proliferation of muscle precursor cells during myogenesis.

However further studies about the mechanisms by which myostatin inhibits muscle growth are needed to better understand the particular muscle development in the hypertrophied Marchigiana cattle breed.

P0482: Cattle

Estimating the Impact of Deleterious Recessive Haplotypes AH1 and AH2 on Reproduction in Ayrshire Cattle

Aline R. Guarini[1], Victoria Kroezen[1], Filippo Miglior[2], Daniela A.L. Lourenco[3], **Luiz F. Brito**[1], Mehdi Sargolzaei[4], John B. Cole[5], Ignacy Misztal[3] and Flavio Schenkel[1], (1)University of Guelph, Guelph, ON, Canada, (2)Canadian Dairy Network & University of Guelph, Guelph, ON, Canada, (3)University of Georgia, Athens, GA, (4)University of Guelph and The Semex Alliance, Unknown, ON, Canada, (5)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD
The effects of two deleterious recessive haplotypes on reproduction were investigated in Canadian Ayrshire cattle. Their phenotypic effects on Stillbirth and 56-day Non-Return Rate (NRR) were calculated by estimating the interaction of service sire carrier-status with maternal grandsire (MGS) carrier-status using the Canadian official genetic evaluation models for those traits. The interaction term included 9 subclasses for the 3 possible conditions of each sire and MGS: haplotype carrier, non-carrier, or not genotyped. A total of 394 carriers and 1,433 non-carriers were available for the AH1 haplotype, while for AH2 haplotype the number of carriers and non-carriers was 313 and 1,543, respectively. The number of matings used in the analysis of Stillbirth was 34,312 and 115,935 for heifers and cows, respectively. For NRR, the number of matings used was 49,479 for heifers and 160,528 for cows. No significant effect of AH1 on 56-day NRR was observed, however, there was an unfavorable effect on Stillbirth rate, which was 2% higher for matings of AH1-carrier sires to dams that have an AH1-carrier sire. This effect was observed for both heifers and cows. AH2 had an unfavorable impact on 56-day NRR, with 5.1% more heifers and 4.0% more cows returning to service, while no significant effect was found on Stillbirth rates. Therefore, the harmful effects of AH1 and AH2 on reproduction traits were validated in the Canadian Ayrshire population, suggesting that breeders should put them into account when making mating decisions.

P0483: Cattle

A Genome Wide Association Study in Holstein Cattle

Beth M Lett, University of Wisconsin Madison, Madison, WI and Brian W. Kirkpatrick, University of Wisconsin-Madison, Madison, WI
Twinning is considered an undesirable trait by the dairy industry and identification of markers associated with twinning rate would facilitate selection for reduced incidence. The primary objective of this study was to conduct a Genome Wide Association Study (GWAS) of twinning rate in US Holstein cattle. Approximately 2.9 million calving records from 2010 to 2016 were obtained from Ag Source and edited for breed (Holstein), records per herd (≥100) and daughters per sire (≥100) leaving 1,444,540 calving records from 658,436 Holstein cows. A generalized linear mixed model considering affects of herd, year, season, parity for each daughter record was run to generate sire means for twinning rate corrected for fixed effects for 2,223 sires. Genotype data obtained from the Council for Dairy Cattle Breeding (CDCB) included 60,670 SNPs for 2,067 of the sires. GWAS analysis was conducted in R using the *GenAbel* package. SNPs with call rate <95%, minor allele frequency <0.01, or Hardy Weinberg Equilibrium test < 0.000001 were excluded. Samples with call rate <95% were likewise excluded. The analysis was run on 58,119 SNPs and 2,061 individuals. False discovery rate (FDR) was calculated using R package *qvalue*. Preliminary results indicated 6 SNPs on chromosomes 5,13,15,22, and 29 that were significant at $P<5x10^{-5}$ and FDR of 0.5 to 0.05. Results will be validated in future work using independent Holstein data.

P0484: Cattle

GWAS and Fine-Mapping of 35 Production, Reproduction and Conformation Traits with Imputed Sequences of 27K Holstein Bulls

Jicai Jiang[1], Paul M. VanRaden[2], John B. Cole[2], Yang Da[3] and Li Ma[1], (1)Department of Animal and Avian Sciences, University of Maryland, College Park, MD, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Department of Animal Science, University of Minnesota, Saint Paul, MN
Fine-mapping of causal variants is becoming feasible for complex traits in livestock GWAS, as an increasing number of animals are sequenced. Imputation has been routinely applied to ascertain sequence variants in large genotyped populations based on small reference populations of sequenced animals. Using the run 5 data of the 1000 Bull Genomes project, we imputed three million sequence variants for 27,000 Holstein bulls after QC edits and LD pruning. These bulls were selected to have highly reliable PTAs for 35 production, reproduction, and body conformation traits. We first performed whole-genome single-marker scan for the 35 traits using the mixed-model based association test in MMAP (https://mmap.github.io). The single-trait association statistics were then merged in multi-trait analyses of three groups of traits, production, reproduction and body conformation, respectively. Two-Mb long candidate genomic regions were selected based on the multi-trait analyses and used in fine-mapping studies. We implemented a state-of-art fine-mapping procedure with a Bayesian method that can assign a posterior probability of causality to each variant and for each independent association signal generate a minimum set of associated variants whose total posterior probability of causality exceeds a threshold (e.g. 95%). Our fine-mapping identified 134 candidate genes for production traits, 97 for reproduction traits, and 156 for body conformation traits, respectively, including some previously reported causal variants, e.g., Chr6:38027010 in *ABCG2* for production traits and Chr7:93244933 in *ARRDC3* for reproduction and body conformation traits. The candidate variant list may facilitate follow-up functional validation and expand our understanding of complex traits in dairy cattle. Additionally, our method can be readily applied to other species where large-scale sequence genotypes are available.

P0485: Cattle

Genetic Effects on Milk Pregnancy Associated Glycoproteins of Holstein Cattle

Daniel Jordan de Abreu Santos, São Paulo State University - UNESP, Jaboticabal, Brazil, John B. Cole, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, Daniel J. Null, Animal Genomics and Improvement Laboratory, ARS-USDA, Beltsville, MD, Todd Martin Byrem, Antel BioSystems, Lasing, MI and Li Ma, Department of Animal and Avian Sciences, University of Maryland, College Park, MD
Pregnancy-associated glycoproteins (PAG) are secreted by the trophoblasts, and are detectable in maternal circulation around the time of attachment of the fetal placenta (3rd week) as well as in blood and milk throughout gestation. While PAG are used as pregnancy biomarkers, they are produced by the trophoblast and, as such, represent one of the earliest phenotypes of the conceptus. The aim of this study is to characterize the genetic effects of milk PAG concentrations by estimating genetic parameters and identifying genomic regions with large

effects on the phenotype. The PAG data used in this study were provided by Antel BioSystems (Lansing, MI), and were combined with animal identification and pedigree information from the National Dairy Database. The analysis included 1,993 pregnancy diagnoses obtained using commercial PAG assays collected from 2012 to 2017. The genotypes data consisted 54,123 SNPs, including autosomal and X chromosomes markers of 2,352 individuals (embryo/fetus and dam). The analysis was performed using a single-step GWAS as implemented in the BLUPF90 family of software (University of Georgia, Athens). The model included contemporary group (herd, year and season) as a fixed effect, embryo/fetus age as covariate, and random direct and maternal additive genetic effects. Estimated h^2 for direct and maternal additive genetic effect were 0.23±0.05 and 0.11±0.05, respectively, and the genetic correlation between these effects was 0.001±0.6. Two genomic regions explained substantial genetic variance for the direct genetic effect: 37,570,928–37,839,397 bp on BTA 19 and 40,312,728–40,626,477 bp on BTA 29, which includes a cluster of bovine PAG genes; and one region had a large maternal effect: 46,919,802–51,441,973 on BTA 4. Six genes had direct effects related to cell processes, and three genes had maternal effects, with *WNT2* being related to placental vasculature. These results suggest that there is moderate genetic control of PAG levels, and may be useful in breeding programs as a measure of fertility.

P0486: Cattle

Identification of Genetic Loci Associated with Conception Rate in US Holstein Dairy Cows

Logan R. A. Hulst[1], Joao Moraes[2], Jennifer Kiser[1], Mahesh Neupane[1], Joseph Dalton[3], Greg Burns[2], Thomas Spencer[2] and Holly L. Neibergs[1], (1)Department of Animal Sciences, Washington State University, Pullman, WA, (2)Division of Animal Sciences, University of Missouri, Columbia, MO, (3)Department of Animal and Veterinary Sciences, University of Idaho, Caldwell, ID

Cow conception rate (CCR), the number of pregnancies that result from dairy cows exposed to a single breeding event, ranges from 30 to 40% in the US, as 40 to 56% of these pregnancies are lost within the first month following conception. Low CCR is in part due to the high level of embryonic loss that occurs after fertilization. The objective of this study was to identify loci associated with CCR in primiparous Holstein dairy cows. Cows (n = 2,015) from six dairies in central Washington with no records of disease or lameness, were bred after observed estrus at 45-50 days post parturition. Pregnancy was determined at day 35 via rectal palpation. A genome-wide association analysis, using an EMMAX-GRM statistical approach with times bred as the phenotype, and a correction for month and year, was conducted with genotypes using the Illumina BovineHD BeadChip on cows that conceived after one (n = 498) or 4 or more breedings (n = 469). Cow conception rate was associated with 74 loci ($P < 5.0 \times 10^{-8}$) of which 46 loci contained a positional candidate gene and 7 loci affected transcription binding factors at the SNP site. Pseudo-heritability was estimated at 0.46. Two of the loci associated with CCR on BTA 4 and 27 were previously reported as associated with heifer conception rate. Loci associated with CCR in primiparous cows will be useful for improving CCR and thus fertility in US Holstein dairy cows through genomic selection.

P0487: Cattle

Inbreeding Measures Using Pedigree and Runs of Homozygosity in Holsteins

Gabriele Marras[1], Filippo Miglior[1], Bayode Makanjuola[1], Flavio Schenkel[2], Mehdi Sargolzaei[3] and Christine F. Baes[1], (1)University of Guelph, Guelph, ON, Canada, (2)Centre for the Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (3)University of Guelph and The Semex Alliance, Unknown, ON, Canada

Abstract: Due to increased selection intensity in dairy cattle, inbreeding levels have steadily increased over the past decades. Genomic information can be analyzed to better capture Mendelian sampling, thereby enabling more accurate estimation of inbreeding. Continuous stretches of homozygous genotypes (runs of homozygosity, ROH) have been shown to provide a better estimate of autozygosity at the genomic level than conventional measures based on pedigree (PED) information alone. In this study, we ed average inbreeding and the variability and rate of increase in ROH-based inbreeding (F_{ROH}) and pedigree-based inbreeding (F_{PED}) using Holstein data collected over the last three decades. A total of 130,727 Holstein animals (66,425 males and 64,302 females) were genotyped using the Illumina BovineSNP50 BeadChip between 1990 and 2017. Runs of homozygosity were calculated using PLINK v1.9 and F_{ROH} was estimated following McQuillan *et. al.* (2008). Our results confirmed that the level of inbreeding calculated as F_{ROH} is considerably higher (0.11 ± 0.03), than that based on F_{PED} (0.06 ± 0.02). Furthermore, the variability of F_{ROH} measures was higher than that of F_{PED}, although generally the rate of increase of inbreeding levels measured using F_{ROH} and F_{PED} was similar (0.07 for F_{ROH} and 0.06 for F_{PED} from 1990 to 2016). Our results also indicated that within the last five years, males have shown a higher increase in F_{ROH} than females, +0.001 in 2011 to +0.028 in 2016, which could be a result of higher selection intensity through genomic selection in males.

P0488: Cattle

Identification of Genetic Markers Associated with Beta Carotene Levels in Buffalo and Dairy Cattle Milk: An Opportunity to Improve Milk Quality in India

Francesca Bertolini[1], Josue Chinchilla Vargas[1], J.R. Khadse[2], Alok Juneja[2], Prasad Deshpande[2], Vinod Potdar[2], Kaustub Bhave[2], A.B. Pande[2] and Max F. Rothschild[1], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)BAIF Development Research Foundation, Pune, India

Beta-carotene, a precursor of vitamin A, can be consumed in buffalo and cattle milk but intake is too low in developing countries. The main genes related to beta-carotene/vitamin A production are *BCMO1*, *BC02* and *SCARB1*. In this work, a Sequenom panel was built using Single Nucleotide Polymorphisms (SNPs) derived by next generation sequencing data from the coding sequence of the three genes. The panel included 44 SNPs for cattle and 23 SNPs for buffalo. A total of 1,421 buffalo (Jaffarbadi, Murrah, Pandharpuri, Mehsana, Surti) and 2,312 cattle (Sahiwal, Tharparkar, Gir, Jersey-cross and Holstein-cross) from India were genotyped with this panel and milk beta-carotene content was measured. The beta-carotene analyses demonstrated a significant difference among cattle and buffalo with buffalo showing a lower beta-carotene content. Among buffalo breeds, the average of beta-carotene levels range from 1.19 mcg/100ml in Mehsana to 4.25 mcg/100ml in Jaffarbadi. For Mehsana and Surti, no suggestive or significant association has been detected. For Pandharpuri, SNPs in both the *SCARB1* and *BCMO1* genes were associated with beta-carotene levels. For Murrah, SNPs in *BCMO1* and *BC02* and for Jaffarbadi, at least one SNP for each gene were associated with beta carotene levels. Several SNPs of the three genes in cattle breeds were associated as well. These markers may be

useful to develop genetic selection strategies that can increase beta-carotene content in milk in Indian animals and the panel may be useful also in other developing countries. Funding for this project was kindly provided by Bill & Melinda Gates Foundation.

P0489: Cattle

Eigen Decomposition Expedites Longitudinal Genome-Wide Association Studies for Milk Production Traits in Chinese Holstein

Chao Ning[1], Dan Wang[1], Xianrui Zheng[1], Qin Zhang[1], Shengli Zhang[1], Raphael Mrode[2] and **Jian-Feng Liu**[1], (1)China Agricultural University, Beijing, China, (2)International Livestock Research Institute, Nairobi, Kenya

Background: Pseudo phenotypes, such as 305-day yields, estimated breeding value (EBV) or deregressed proof (DRP), are usually utilized as response variables of genome-wide association study (GWAS) for milk production traits of dairy cattle. Computational inefficiency challenges direct utilization of test-day records for longitudinal GWAS in large data.

Results: We proposed a rapid longitudinal GWAS method based on random regression model. Our method utilized eigen decomposition of phenotypic covariance matrix to rotate the data, thereby transforming the complex mixed linear model into weighted least squares. We applied our method to the analysis of milk production traits of first three parities in Chinese Holstein. The analysis for each trait could be finished within a day with known variances. Furthermore, the genomic inflation factors were much closed to the desired level of one. In total, we located 79 significant single nucleotide polymorphisms (SNPs) and 72 of them were within reported quantitative trait loci (QTL) region.

Conclusions: Our rapid method can control Type I error in longitudinal data analysis, and can be applied to other longitudinal traits. The results in Chinese Holstein validate those from previous study. Moreover, six additional SNPs for fat percentage and one for protein percentage were located by our method. These seven SNPs can be new candidate quantitative trait nucleotides (QTNs) for milk production traits in Chinese Holstein.

Acknowledgments: We appreciate the financial support from the National Major Development Program of Transgenic Breeding (2014ZX0800953B).

P0490: Cattle

Copy Number Variation in Dairy Cattle using Next-Generation Sequencing

Tatiane C. S. Chud[1], Derek M. Bickhart[2], Adhemar Zerlotini[3], John B. Cole[4], Marcos Vinicius B. da Silva[5] and Danísio P. Munari[1], (1)Universidade Estadual Paulista (Unesp), Jaboticabal, São Paulo, Brazil, (2)Dairy Forage Research Center, USDA-ARS, Madison, WI, (3)Embrapa Informática Agropecuária, Campinas, SP, Brazil, (4)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (5)EMBRAPA Dairy Cattle – MG, Juiz de Fora, Brazil

Gene copy number variants (CNV) have been shown to be associated with several production traits in dairy cattle; however, the detection and validation of CNVs in crossbred cattle is currently lacking. In order to provide a basis for future association studies, we sought to identify CNV regions (CNVRs) within the Girolando composite breed resulting from a mating of the Holstein (taurine) and Gir (indicine) breeds. A read depth method was performed using CNVnator software on NGS data from two Girolando, two Gir and ten Holstein bulls. The individual CNVs were merged into CNVRs based on genomic regions overlapping by at least 1 bp. In total, we identified a composite of 1,286 CNVRs (520 deletions, 255 duplications, 511 mixed) on the genomes of all samples. We observed 34 CNVRs (nine deletions, 25 mixed) in common (overlapping > 50%) only between Girolando and Holstein and 181 CNVRs (20 deletions, 21 duplications,140 mixed) only in Girolando and Gir, suggesting parent-of-origin inheritance from Holstein and Gir cattle, respectively. One of these Holstein-specific CNVRs intersected with the interleukin 6 family cytokine (*LIF*) gene which is linked to fat production and fertility traits in Holstein. Genes related to disease resistance (e.g. the *CD4* gene) also coincided with CNVRs present only in Gir and Girolando cattle suggesting an indicine origin for the CNV. These results showed evidence of specific CNVRs shared by Girolando and purebred breeds which may be targeted for future selective breeding.
Financial Support: FAPESP fellow 2015/08939-0.

P0491: Cattle

Genome-Wide Scan of Copy Number Variants (CNVs) in Valdostana Red Pied and Comparison with Italian Brown Swiss and Mexican Holstein

Erica Gorla[1], Maria C. Cozzi[1], Mario Vevey[2], Maria Longeri[1], Francesca Genova[1], Sergio I. Roman-Ponce[3], Felipe de Jesús Ruiz[3], Marina Duran-Aguilar[4], Alessandro Bagnato[1] and Maria G. Strillacci[1], (1)Università degli Studi di Milano - Dept. DiMeVet, Milano, Italy, (2)Associazione Nazionale Allevatori Razza Valdostana, Gressan, Italy, (3)Centro Nacional de Investigación en Fisiología y Mejoramiento Animal INIFAP, Ajuchitlán, Queretaro, Mexico, (4)Universidad Nacional Autonoma de Mexico, Queretaro, Mexico

Copy Number Variants (CNVs) are an important source of genomic structural variation. Many studies have focused on identified CNVs within and between populations, in livestock as well as in human, but only a few studies have explored population-genetic proprieties in cattle based on CNVs derived from the high-density SNP array. Here we report a high resolution CNV scan from Illumina's 777k BovineHD Beadchip data for the Valdostana Red Pied (VRP), an autochthonous Italian dual-purpose cattle population reared in the Alps that did not undergo strong selection for production traits. We perform a genetic comparison among the VPR, the Italian Brown Swiss (IBS 164 sires) and the Mexican Holstein (HOL 124 males and females) based on CNVs. In the VRP we identified a total of 6,784 CNVs that were summarized to 1,723 CNV regions (CNVR) on 29 autosomes covering a total of ~59 Mb of the UMD3.1 autosome. A total of 171 CNVRs were shared by all breeds, 474 were common to VRP and IBS, while 313 overlapped between VRP and HOL, indicating a more similar genetic background for the populations originating from the Alps. The PCA and the NJ tree showed a clear separation of the two Alpine breeds from the HOL. Vst statistics was calculated for CNV common regions among breeds identified in at least 5 individuals. Genes annotated in the CNVRs were compared to disclose diverging paths of selection among breeds.

P0492: Cattle

Copy Number Variations Mapping in Mexican Criollo Population

Maria C. Cozzi[1], Erica Gorla[1], Sergio I. Roman-Ponce[2], Vicente Murillo Vega Eliezer[3], Felipe de Jesús Ruiz[2], Marina Duran-Aguilar[4], Moisés M Montaño-Bermúdez[2], Guillermo Martinez Velasquez[5] and Maria G. Strillacci[1], (1)Università degli Studi di Milano - Dept. DiMeVet, Milano, Italy, (2)Centro Nacional de Investigación en Fisiología y Mejoramiento Animal INIFAP, Ajuchitlán, Queretaro, Mexico, (3)Campo Experimental La Posta, INIFAP, Medellín, Mexico, (4)Universidad Nacional Autonoma de Mexico, Queretaro, Mexico, (5)INIFAP-Campo Experimental Santiago Ixcuintla, Santiago Ixcuintla,, Mexico
The Mexican Criollo Cattle (MCC) derives from cattle of the Iberian Peninsula, mainly from Andalucía and Extremadura, and was introduced in Central America the 16th century. Copy Number Variations (CNVs) are large-scale structural DNA insertions and deletions, ranging from 50bp to several megabases (Mb) strongly contributing to the genetic diversity and to phenotypic expression. More than half of the identified CNVs regions in human and animals affect protein-coding genes expression especially related to disease susceptibility. The aim of this study is to map the CNVs in the MCC population using two different algorithms. A total of 48 unrelated individuals (5 males and 43 females) were genotyped with the Illumina BovineHD Genotyping BeadChip, containing 777,962 polymorphic SNPs. After stringent quality control, CNVs calling was performed with the CNAM of Golden Helix SVS 8.3.1 (SVS) and with the Hidden Markov Model of PennCNV software. CNVs were summarized into CNV regions (CNVRs) at a population level, using BEDTools.
The SVS software identified 740 CNVs in 43 individuals on 27 autosomes. Among these, 739 were deletions and only one was classified as duplication. The number of CNVRs identified was 209. The PennCNV software detected 1,849 CNVs (1,206 deletions and 643 duplications) in 43 individuals on 29 autosomes, summarized to 1,013 CNVRs (261 gains, 698 losses and 54 complex). The obtained data were compared with the results previously obtained in other populations with the same algorithms. The consensus map using the CNVRs from the two software permitted to reduce the false positive calling.

P0493: Cattle

Comparative Genome Wide Characterisation of Five Rare British Isles Cattle Breeds

Paul Flynn, Weatherbys Scientific, Naas, Ireland
Several British Isles cattle breeds have been subjected to population reductions over the past few centuries. Five such breeds are Kerry (KY), Dexter (DX), DroimFhionn (DF), Irish Moiled (IM) and White Park (WP). Comparative genome wide characterisation studies contribute towards conservation strategies to ensure future survival of such Bovine genetic resources.
Using a 12,890 SNP subset from IDB SNP chip, this study established genetic parameters such as diversity, differentiation and population structure. Samples were collected from each breed (total n=225), with selection based on pedigree knowledge to maximise breed representation.
Reduced levels of breed genetic diversity along with genetic isolation and strong population structure were observed for WP (He 0.36502 ± 0.14465) and IM (He 0.36712 ± 0.14396). Greatest genetic distance was observed between WP and IM (Fst 0.21624, P<0.01). KY (He 0.41602 ± 0.10763) and DX (He 0.43498 ± 0.08935) displayed comparative genetic diversities, however Principle Component and Structure analysis generated distinct clusters between breeds. DF breed displayed genetic diversity (He 0.41854 ± 0.10712) similar to overall mean (He 0.41214 ± 0.10882) with comparative genetic distances ranging from closest – Holstein Friesian (Fst 0.09263, P<0.01) to furthest - WP (Fst 0.17342, P<0.01).
Data set findings included reduced genetic diversity and genetic isolation for IM and WP, along with distinct differentiation between KY and DX. Novel insights into DF genetic parameters were revealed within this breed's first genome analysis. Results yielded a genome wide snapshot of current genetic status for each breed and provides a platform to implement future breeding strategies.

P0494: Cattle

Chromosome Level Assembly of Tibetan Yak Using Synteny to Cattle

Samathmika Ravi, Institute of Bioinformatics and Applied Biotechnology, Bengaluru, Karnataka, India, **Dr. Shubham Dixit**, Institute of Bioinformatics and applied biotechnology, Bangalore, India and Subhashini Srinivasan, Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India
In today's NGS driven era, creating a reference genome for every organism of economic interest is within reach. However, complexity of eukaryotic genomes and limitations of sequencing/assembly processes slacken generation of reference genomes for use in breeding. Although state-of-the-art complementing technologies like Hi-C and BioNano enable us to generate assembly at chromosome level resolution, they come at a huge cost. However, as reference genomes of many species are assembled, draft genomes of closely related species can be assembled using shared synteny. Here, we present our effort in attaining a chromosomal level assembly for Tibetan yak using a two step method. Firstly, improving the existing draft assembly of yak by assisted assembly approach using the reference genome of cattle. We have been able to place 95% of the improved yak assembly onto cattle chromosomes using the synteny between the two species. The final assembly of the 30 chromosomes of yak has been attained by placing yak scaffolds onto cattle reference using an identity of greater than 90% at the protein level as cutoff. The N50 of the yak genome has improved to 102.95 Mbp from 25.25 Mbp. This resource for yak is fundamental to gaining further insights into breeding, domestication and its evolutionary closeness to other bovid species.

P0495: Cattle

Genomic Prediction of Buffalo Milk Traits by Incorporating the Related Cattle QTLs

Xingjie Hao[1,2], Xiang Zhou[2], Aixin Liang[1], Graham S Plastow[3], Zhiquan Wang[3], Chunyan Zhang[3], JiaJia Liu[1], Giuseppe Campanile[4], Gasparrini Bianca[4], Angela Salzano[4], Shujun Zhang[1] and Liguo Yang[1], (1)Huazhong Agricultural University, Wuhan, China, (2)University of Michigan, Ann Arbor, MI, (3)University of Alberta, Edmonton, AB, Canada, (4)University of Naples Federico II, Naples, Italy
The 90K Axiom Buffalo SNP Array is expected to improve and speed up various genomic analyses for the buffalo (*Bubalus bubalis*). Genomic prediction is an effective approach in animal breeding to improve selection and reduce costs. While buffalo genome research is lagging behind that of the cow, and production records are also limited, genomic prediction performance will be relatively poor. We introduced a new approach (pGBLUP) for genomic prediction of six buffalo milk traits by incorporating QTL information for cattle milk traits in order to help

improve the prediction performance in buffalo. In simulations, pGBLUP could outperform BayesR and GBLUP if the prior biological information was ideal; otherwise it performed slightly worse than BayesR, and equal or better than GBLUP. In real data, the heritability of buffalo genomic region corresponding to the cattle milk traits QTLs was enriched in four buffalo milk traits (FY270, MY270, PY270 and PM) when EBV as the response variable. DEBV as the response variable yielded more reliable genomic predictions than traditional EBV as has been shown by previous research. To our knowledge, this study was the first to apply genomic prediction in buffalo by incorporating prior biological information. The genomic prediction for buffalo traits can be further improved with larger sample size, higher density SNP chips and more precise prior biological information.

P0496: Cattle

Convergent Evolution of Rumen Microbiomes in High-Altitude Mammals

Dongming Xu, Kunming Institute of Zoology, Kunming, AK, China and Zhigang Zhang, KIZ, CAS, Kunming, China

Studies of genetic adaptation, a central focus of evolutionary biology, most often focus on the host's genome and only rarely on its co-evolved microbiome. The Qinghai-Tibetan Plateau (QTP) offers one of the most extreme environments for the survival of human and other mammalian species. Yaks (Bos grunniens) and Tibetan sheep (T-sheep) (Ovis aries) have adaptations for living in this harsh high-altitude environment, where nomadic Tibetan people keep them primarily for food and livelihood. Adaptive evolution affects energy-metabolism-related genes in a way that helps these ruminants live at high altitude. Herein, we report convergent evolution of rumen microbiomes for energy harvesting persistence in two typical high-altitude ruminants, yaks and T-sheep. Both ruminants yield significantly lower levels of methane and higher yields of volatile fatty acids (VFAs) than their low-altitude relatives, cattle (Bos taurus) and ordinary sheep (Ovis aries). Ultra-deep metagenomic sequencing reveals significant enrichment in VFA-yielding pathways of rumen microbial genes in high-altitude ruminants, whereas methanogenesis pathways show enrichment in the cattle metagenome. Analyses of RNA transcriptomes reveal significant upregulation in 36 genes associated with VFA transport and absorption in the ruminal epithelium of high-altitude ruminants. Our study provides novel insights into the contributions of microbiomes to adaptive evolution in mammals and sheds light on the biological control of greenhouse gas emissions from livestock enteric fermentation.

P0497: Poultry

Gene Regulatory Mechanisms in Chicken Primordial Germ Cells Directing the Formation of Male and Female Gametes

Dadakhalandar Doddamani[1], Mark Woodcock[1], Lorna Taylor[1], Jacqueline Smith[2], David W. Burt[1] and Mike McGrew[1], (1)The Roslin Institute, The University of Edinburgh, Edinburgh, United Kingdom, (2)The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom

Primordial germ cells (PGCs) are the cells central to germline competency as they give rise to the functional gametes of the animal. The potential usage of PGCs in producing gametes in vitro has driven research in the derivation, culturing and manipulations of PGCs. In chicken, these cells appear at an early stage of embryo development followed by migration to the germinal ridges. After proliferation for several days in the gonads, they develop a gender-specific developmental pathway that will lead them to become either egg or sperm. To better understand the genetic mechanisms controlling the choice between the formation of a male and female gamete, we aim to study differences in gene expression, gene regulatory networks and cellular pathways between male and female chicken PGCs. An average of ~119 million Illumina paired-end reads of length 150 base pairs was obtained and used to construct the transcriptome for each sample. The transcriptome reveals germ cell-specific genes, genes involved in epigenetic reprogramming and regulation of cell cycles are constitutively expressed in both male and female PGCs. We identified 107 and 256 genes are differentially expressed in male and female PGCs respectively. These genes will further investigated to identify the regulatory network involved in the formation of functional gametes. The current study will also help in identifying the mechanisms controlling the sex-specific in vitro growth requirements of germ cells and will be used in our future avian PGC research.

P0498: Poultry

Genetic Variations of Egg Quality in Chickens at Late Laying Period Revealed By Genome-Wide Association Study

Zhuang Liu, China Agricultural University, Beijing, China

With the extension of laying cycle, the rapid decline of egg quality at late laying period has aroused great concern for the poultry industries. To solve this problem, many researches had focused on the nutrition of chicken at the end of laying cycle, but the effect was limited. Herein, we performed a genome-wide association study (GWAS) to identify genomic variations associated with egg quality employing the chicken 600K high density SNP array in a population of 1078 hens at 72 and 80 weeks of age. Results indicated that a genomic region spanning from 8.95 to 9.31 Mb (~0.36Mb) in GGA13 was detected to be significantly associated with albumen height (AH) and haugh unit (HU) and the two most significant SNPs could accounted 3.12 ~ 5.75% phenotypic variance. Two promising genes including MSX2 and DRD1 were mapped to the narrow region, which was involved in embryonic and ovary development and found to be related to egg production, respectively. Moreover, three interesting genes, RHOA, SDF4 and TNFRSF4 identified from three missense loci were considered to be candidate genes regulating egg shell color. Findings in our study could provide worthy theoretical basis and technology support to improve late-stage egg quality for breeders.

Key words:egg quality; SNP; GWAS; candidate genes

P0499: Poultry

Proteomics, Transcriptomics and Functional Studies of Candidate Genes for Controlling Biomineralization Process in Birds Eggshell

Ya Hui Gao, Quan Zhang, Feng Zhu, Zhonghua Ning and Zhuocheng Hou, China Agricultural University, Beijing, China

Eggshell damages lead to economic losses in the egg production industry and result in microorganisms contamination threating to human health. In recent years, we have characterized the transcriptome, whole genome sequencing and proteomics for identifying the key genes and genetic mutations associated with eggshell calcification in birds. Using multiple Omics data and previous results help us to probe several functional genes. OC-17 is one of the best candidates to control and regulate the deposition of calcium carbonate in the calcified eggshell layer. To examine the special function of protein OC17 on the calcium carbonate precipitation and identify the main functional residues of protein

OC17, recombinant expression of wild and mutant OC17 proteins in mammalian cells of HEK 293 T were conducted firstly. To construct mutant proteins, all of the candidate functional residues which were determined by sequence alignment with C-type lectin-like proteins from vertebrate calcitic biominerals and molecular dynamics simulations were replaced with alanine. The target proteins were secreted into the culture medium and collected post-transfection 3 d. The yield of target proteins was 2 mg per litter after purification using immobilized-metal affinity chromatography method. The results of biomineralization assay in vitro demonstrated that both wild type and mutated (serine-61 deletion) forms of recombinant OC17 have a significant but similar impact on calcium carbonate crystals. Both mutants can inhibit the formation of calcite <104> face and delay the growth of crystal, which is strikingly different from calcium carbonate precipitation patterns in the absence of protein.

P0500: Poultry

Evaluation of Hansen Probiotics for Protection against BCO Lameness on Litter with *Staphylococcus agnetis* Challenge

Amer Hasan, University of Arkansas, Fayetteville, AR

Lameness is one of the main metabolic diseases related to fast growth in broilers. It is a significant problem in the poultry industry resulting in hundreds of millions of dollars in lost revenue annually. In commercial broilers, the most common cause of lameness is bacterial chondronecrosis with osteomyelitis (BCO). BCO has been associated with a broad range of bacterial species, where Staphylococcus species are a major bacterial isolated from BCO lesions. We have characterized a hypervirulent *Staphylococcus agnetis* isolate (908) as the major species isolated from the lame broilers in our research farm. We have never identified another BCO isolate as virulent in our *in vivo* or *in vitro* assays. Previously, we had identified particular probiotics that can reduce BCO lameness. In this experiment, we tested a standard diet supplemented with probiotics from Chr. Hansen Holding A/S (Denmark). Birds on control and probiotic diets were challenged with *S. agnetis* 908 on days 20 and 21. The birds on the diet formulations with either GalliPro Tect or GalliPro Max were significantly reduced for lameness compared to the birds on the control diet. Severity of Femoral and Tibial BCO lesions in lame birds were also reduced with the two probiotics. Cumulative lameness at day 56 was 77% for the control diet with bacterial challenge, 63% for unchallenged birds on control diet, and 48-50% for challenged birds on either of the two probiotic supplemented feeds. Thus, the probiotics reduced lameness by 35-37%, despite a challenge with a hypervirulent strain. That reduction in lameness is significant and greater than any other intervention we have found to date.

P0501: Poultry

Investigation for High-Impact SNPs on Important Genes for Muscle Growth Characteristics in Broilers

Bruna Petry[1], Gabriel C. M. Moreira[2], Erika Cristina Jorge[3], Monica Correa Ledur[4] and Luiz L. Coutinho[2], (1)University of São Paulo - Brazil, Piracicaba, NB, Brazil, (2)University of São Paulo, Piracicaba, Brazil, (3)University of Minas Gerais, Belo Horizonte, Brazil, (4)Embrapa Swine and Poultry, Concordia, Brazil

The study of candidate genes and mutations is necessary to understand the biological processes and the genetic networks that control phenotypic traits of economic importance. The aim of this study was to find high-impact mutations in candidate genes for muscle growth and development traits in a Brazilian meat-type chicken population. For this, we used a dataset of 11 million SNPs found in 14 parental chickens of this population. After functional annotation, the SNPs predicted as high-impact and also annotated within QTLs for breast meat weight and percentage traits were kept. From the genes harboring these SNPs we selected those related with muscle development traits based on their gene ontologies. We identified important genes related to processes of cell division and proliferation. As an example, one of the QTLs previsiouly identified by our group in this population was on GGA 26 (2 Mb) associated with breast meat weight. In this window, three important genes related to muscle development were found: *TRIM33,* regulating the transforming growth factors beta, *RHOC,* related with migration of satellite skeletal muscle cells and *OLFML3,* an important gene acting in multicellular development. All these genes exhibited high-impact SNPs and, therefore, are important candidates for muscular growth and development, cell division and proliferation. Further studies aiming to validate those variants and their effects are important for a better understanding of the molecular mechanisms involved in muscle development in chickens.

P0502: Poultry

Inherited CNVs in Regions Related to Breast Muscle Development in a Broiler Population

Thais Fernanda Godoy[1], Vinicius H. Silva[2], Gabriel C. M. Moreira[1], Clarissa Boschiero[3], Monica Correa Ledur[4], Richard Crooijmans[5], Martien A.M. Groenen[6] and Luiz L. Coutinho[1], (1)University of São Paulo, Piracicaba, Brazil, (2)Wageningen University, Wageningen, Netherlands, (3)The Noble Research Institute, Ardmore, OK, (4)Embrapa Swine and Poultry, Concordia, Brazil, (5)Wageningen University & Research, Wageningen, Netherlands, (6)Wageningen University & Research Center, Wageningen, Netherlands

Copy number variation (CNV) is an important genome polymorphism that contributes to phenotypic variation in a wide range of species, including chicken. We studied inherited CNVs in a Brazilian broiler population that overlap with breast muscle weight QTLs. A total of 826 chickens developed by Embrapa were genotyped using an Affymetrix 600K SNP array. The allele frequencies and intensities were analyzed by PennCNV (joint option) which detected 21,077 inherited CNVs. The CNVs were concatenated into 3,240 CNV regions (CNVRs) which cover 128.46 Mb (10.44%) of the autosomal chicken chromosomes (Galgal5). The length of the CNVRs ranges from 1 to 2,887 kb (average of 39.65 kb) and includes 492 losses, 2,461 gains and 287 harboring both losses and gains. Of these CNVRs, 41 overlap with breast muscle weight QTLs. In these regions, the CNVRs overlap with 79 genes. A CNVR gain of 20,771 bp was observed in *FGFR2,* a gene related with growth of fibroblasts. Two other CNVRs (a gain of 1,212 bp and a loss/gain of 50,942 bp) overlap with *WNT9A* and *SOCS3* respectively, two genes involved in cell differentiation. A gain CNVR of 6,414 bp overlaps with the *SLIT2* gene while another gain CNVR of 2,458 bp overlaps with *HTRA1*, both genes that are related with cell growth. In future studies, genome-wide association analysis with CNVs for breast muscle traits will complement the current results, helping to understand the genetic architecture of breast muscle development in chickens.

P0503: Poultry

Single Nucleotide Polymorphism Associated with Breast Trait in Chickens

Priscila Anchieta Trevisoli[1], Gabriel C. M. Moreira[1], Clarissa Boschiero[2], Juliana Petrini[1], Gerson B. Mourão[3], Monica Correa Ledur[4] and Luiz L. Coutinho[1], (1)University of São Paulo, Piracicaba, Brazil, (2)The Noble Research Institute, Ardmore, OK, (3)ESALQ - USP, Piracicaba, SP, Brazil, (4)Embrapa Swine and Poultry, Concordia, Brazil

In previous studies conducted in our research group, two QTLs were mapped for breast weight (BRW) and yield (BR%) in a paternal broiler male line (TT). Within these QTLs, analysis of resequencing data from 14 parental animals resulted in the identification of 12 non-tolerated Single Nucleotide Polymorphisms (SNPs), determined by SIFT score, in nine candidate genes. For each SNP, a region of 75 bp up and down stream was amplified and sequenced in 237 animals with Illumina Custom Amplicon Sequencing technology. The raw sequencing data was aligned against the chicken reference genome (Galgal_5.0) and the SNP calling was performed with SAMtools software followed by functional annotation using VEP tool. Association analysis for BRW and BR% with the non-tolerant SNPs were performed one SNP at time, with the Linear Mixed Model in SAS software. Body weight at 42 days of age was used as a covariate for BRW; sex, hatch and SNP were considered as fixed effects, family and residual as random effects. Only one SNP located within the *WDR77* gene (GGA26, rs736010549) was associated ($p<0.1$) with BRW. This gene is related to the activation or increase in the rate or extent of cell proliferation and differentiation and has an important role in protein-protein complexes formations. Our results are not conclusive. We can not differentiate if rs736010549 is the causal mutation or if it is in linkage disequilibrium (LD) with the causal mutation. Further functional studies are necessary to elucidate these findings.

P0504: Poultry

Allele Specific Expression in Chicken Liver, Muscle and Adipose Tissues

M. Joseph Tomlinson IV[1], Shawn Polson[1], William Lee[2] and Behnam Abasht[1], (1)University of Delaware, Newark, DE, (2)Maple Leaf Farms, Leesburg, IN

Allele specific expression (ASE) is the process where one allele in a heterozygous individual is expressed at a higher level in comparison to the other when equal expression is expected. ASE is of great interest because it helps in the identification of cis-regulatory mutations or epigenetic modifications that influence gene expression. In this study, the effect of these cis-regulatory elements is investigated by examining SNPs from RNAseq. Identified SNPs may then be utilized in future animal breeding programs. In our study, we performed RNA-sequencing on 100 samples collected from various populations of chickens. We then followed GATK's "Best Practices for Variant Calling on RNAseq" using recommended settings. We aligned the sequence reads to the chicken reference genome sequence (Gallus_gallus-5.0) from Ensemble. Based on the first round of alignment of STAR, the average number of reads was 36,196,012 with an average mapping rate of 85%. We identified a total 3,147,284 variants (SNPs and Indels) and then used these variants to mask the reference genome for initial alignment and re-ran the pipeline. The final variants from a large sub-set of the prior samples (n= 68) were filtered and examined for ASE using the binomial test (p-value=0.01). This subset consisted of samples from different tissues (breast muscle, abdominal fat, liver), but from the same population. On average ~325,000 SNPs in each tissue passed our filter criteria, ~12% of which showing ASE. The overlap of ASE SNPs among the 3 tissues was only ~11%, suggesting that ASE is a tissue dependent mechanism in chickens.

P0505: Poultry

Reducing Campylobacter Colonization, Virulence Factors and Modulating Cecal Microbiome Profile in Broiler Chickens through in-Water Supplementation of Trans-Cinnamaldehyde Nanoemulsions

Abhinav Upadhyay[1], Indu Upadhyaya[1], Komala Arsi[1], **Ann M. Donoghue**[2] and Dan J. Donoghue[1], (1)University of Arkansas, Fayetteville, AR, (2)USDA ARS, Fayetteville, AR

Campylobacter jejuni is a major foodborne pathogen that causes diarrhea in humans. Chickens act as the host for *C. jejuni*, wherein the pathogen colonizes the ceca thereby leading to contamination of the carcass during slaughter and subsequent human infections.
This study investigated the efficacy of in-water supplementation of trans-cinnamaldehyde nanoemulsion in reducing *C. jejuni* cecal colonization in 14-day-old broilers. In addition, the effect of trans-cinnamaldehyde on *C. jejuni* virulence attributes (motility, attachment to chicken enterocytes) and cecal microbiome was investigated.
In two separate trials, day of hatch broiler chickens (Cobb; 10 birds/treatment/trial) were supplemented with trans-cinnamaldehyde (normal or nanoemulsion form) in drinking water at 0, 0.0625, 0.125, 0.25, 0.5, and 1% level. On day 7, the birds were orally challenged with a four-strain cocktail of *C. jejuni* (~6 log CFU/bird). On day 14, the birds were sacrificed and *C. jejuni* colonization in cecal contents were quantified. Administration of 0.25% trans-cinnamaldehyde nanoemulsion (size ~100-200 nm; zeta potential ~ -0.35 mV) reduced *C. jejuni* colonization by ~1.5-2 logs CFU/mL ($P<0.05$). trans-cinnamaldehyde also reduced pathogen motility and attachment to chicken enterocytes ($P<0.05$). No reduction in feed/water consumption or weight gain was observed in 0.25% trans-cinnamaldehyde treatments as compared to controls ($P>0.05$). Illumina MIseq based microbiome analysis revealed significant differences across treatments in the beta diversity of cecal microbiota. The administration of trans-cinnamaldehyde nanoemulsion increased populations of Clostridia as compared to control and other treatments ($P<0.05$). Results suggest that trans-cinnamaldehyde nanoemulsion could potentially be used to control *C. jejuni* colonization in broiler chickens.

P0506: Poultry

RNA Sequencing of Macrophages from Different B-Haplotype Chickens

Yvonne Drechsler, Western University of Health Sciences, Pomona, CA and Kristopher Irizarry, The Applied Genomics Center, College of Veterinary Medicine, Western University of Health Sciences, Pomona, CA

Developing advanced strategies for controlling viral and bacterial infections in different species requires understanding of the underlying mechanisms of disease resistance and susceptibility, including variations in gene expression of immune responses . Macrophages, important cells of innate immunity, are directly involved in cellular interactions with pathogens. This study characterizes the molecular basis for dramatically different activation of macrophages by IFNγ from the B2 and the B19 B-haplotypes. A large-scale RNA sequencing approach was employed to sequence the RNA of differentiating macrophages before and after stimulation. Our results demonstrate that a large number of genes exhibit divergent expression between B2 and B19 haplotype at critical timepoints of differentiation and stimulation. Several immune pathways such (TLR, apoptosis, cytokine-cytokine) show more robust gene expression in the disease resistant birds. In addition, gene

expression of microRNAs and transcription factors show differences in macrophages from B2 versus B19 haplotypes. These differences in gene expression appear to be regulated by complex epigenetic mechanisms that need further investigation.

P0507: Poultry

Comparative Temporal Analysis of Liver in Post-Hatch Broiler Chicks

Heidi A Van Every[1], Bridget A Aylward[1], Ryan Arsenault[1] and Carl J. Schmidt[2], (1)University of Delaware, Newark, DE, (2)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

The modern broiler chicken grows at a rapid rate, reaching market age in only six weeks. Transcriptome, kinome, and metabolome data from the liver can be used collectively to identify key differences during the early shift from lipid to carbohydrate metabolism, and subsequent accelerated muscle growth. For this study Ross708 broilers were grown and necropsied at ten time points between days 2-20 post-hatch. RNAseq data was obtained from liver, and metabolome data from liver and plasma. Kinome arrays were performed on day 6 and 16 liver. Here we focus on days 6 and 16 post-hatch, with an emphasis on transcriptome-predicted kinase activity and observed phosphorylation targets.

P0508: Poultry

Hepatic Transcriptome Analysis of the Chicken Embryo Exposed to Aflatoxin B1

Sung-Jin Lee[1], So-Young Choi[2], Song-Yi Yang[3] and Min-Wook Hong[1], (1)Kangwon National University, Chuncheon, South Korea, (2)Kangwon National Uinv. Korea, Chuncheon, Korea, Republic of (South), (3)Kangwon national university, South Korea, Korea, Republic of (South)

Aflatoxin B1 (AFB1) is highly hazardous mycotoxin in human and animals. Despite the fact that the poultry is especially sensitive to aflatoxicosis, the previous study about the transcriptome has not been reported in chicken eggs exposed to AFB1. The objective of this study was to investigate the global gene expression responses to AFB1 in the embryonic liver. Eggs (n=5) were injected with AFB1 (2 mg and 2 ng/egg) or vehicle and dissected for hepatic tissue. Illumina RNA-sequencing was used to examine transcriptome responses to AFB1 exposure in the chicken embryo. We found 41 and 19 transcripts showing 2-fold or greater differentially expressed genes (DEGs) between the control and 2-dose AFB1 treatments (2 mg and 2 ng/egg), respectively. Specifically, compared to the control, 21 transcripts were up-regulated (> 2-fold) and 18 transcripts were down-regulated (> 2-fold) in treatment with 2mg/egg AFB1. RNA-sequencing is a powerful approach to study global gene expression changes exposed to toxin. These results as the fundamental information about the genes involved in AFB1-induced hepatotoxicosis could help to understand the aflatoxicosis. In addition, it should be analyzed gene ontology about the DEGs and proteome to identify the molecular mechanisms of toxicity in further study.

P0509: Poultry

Modification of ATAC-Seq Permits Profiling of Open Chromatin in Cryopreserved Chicken Lung

Michelle M. Halstead[1], Hans H. Cheng[2], Perot Saelao[1], Kelly Chanthavixay[1], Ying Wang[1], Mary E. Delany[1], Huaijun Zhou[1] and Pablo J. Ross[1], (1)Animal Science, University of California, Davis, CA, (2)USDA, ARS, ADOL, East Lansing, MI

The diversity of cell types found in multicellular organisms stems from differences in their epigenetic state. Chromatin organization and transcription factor binding determine the impact of regulatory elements on transcription, producing cell-specific expression profiles. Regions of 'open' chromatin facilitate DNA-protein interactions, and thereby the function of regulatory elements. While techniques such as DNase I hypersensitive sites sequencing (DNase-seq) are commonly used to profile open chromatin, the Assay for Transposase-Accessible Chromatin (ATAC-seq) has emerged as an alternative due to its simplicity and low cell input requirements.

To apply this technique to samples stored long-term, the ATAC-seq protocol was modified to be compatible with cryopreserved nuclei preparations from tissues. Following these modifications, comparison of sequence data from ATAC-seq and DNase-seq (Stam Lab, University of Washington) on chicken lung demonstrate that 1) the homogenization of tissue, quality of cells, and successful lysis of cells are the most important criteria for good quality data, 2) a library's nucleosomal laddering pattern is not a consistent indicator of library quality, and 3) size selection of libraries for sub-nucleosomal length fragments significantly improved the signal-to-noise ratio after mapping. Following optimization, regions enriched for ATAC-seq reads captured more DNase I hypersensitive sites and regions with active histone modifications; additionally, genes with open chromatin, identified by ATAC-seq, were significantly more likely to be expressed than those without ($p = 2.51e-19$). These results broaden the applicability of ATAC-seq to cryopreserved nuclei preparations from tissues, which will benefit current efforts to annotate regulatory elements in different species.

P0510: Poultry

Variations in Genomic Organization, Gene Structure and Expression Among Chicken MHC-*Y* Haplotypes

Jibin Zhang, Ronald M. Goto and Marcia M. Miller, Beckman Research Institute, City of Hope, Duarte, CA

It is becoming increasingly apparent that the MHC-*Y* region of chicken chromosome 16 is highly variable. We are investigating the role of MHC-*Y* in resistance to infection and colonization by bacteria. Multiple differences exist between MHC-*Y* haplotypes in gene copy number and apparently in expression of genes in response to immunogens including microbes. These differences may be important in how individual chickens with different MHC-*Y* haplotypes respond to colonization and to infectious disease. Nearly finished sequence determinations for MHC-*Y* in Red Jungle Fowl provide a window into the tremendous complexity of MHC-*Y* gene organization.

To begin to better define the genetic basis for differences observed in response to microbes, we are investigating MHC-*Y* haplotypes present among selected lines. We are also examining the influence of haplotype differences on the expression of MHC-*Y* class I genes in simple *in vitro* microbial challenge assays. In these qPCR-based assays, *YF* gene expression is highly sensitive to the presence of microbes, with expression quickly increasing in response to small numbers of bacteria, but with decreasing expression over the same interval when higher numbers of bacteria are added. Longer incubation times result in decreased *YF* gene expression. These early results suggest a role for YF in defining very early immune responses to bacteria.

P0511: Poultry

Detection of Selection Signatures Among Brazilian, Sri Lankan, and Egyptian Chicken Populations Under Different Environmental Conditions

Muhammed Walugembe[1], Francesca Bertolini[2], Mahinda C. B. Dematawewa[3], Ahmed R. Elbeltagy[4], Carl J. Schmidt[5], Susan J. Lamont[1] and Max F. Rothschild[2], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Department of Animal Science, Iowa State University, Ames, IA, (3)University of Peradeniya, Peradeniya, Sri Lanka, (4)Animal Production Research Institute, Dokki, Cairo, Egypt, (5)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Extreme environmental conditions are a major challenge in livestock production. Changes in climate, particularly those that contribute to weather extremes like drought or excessive humidity may result in reduced performance and reproduction and could comprise the animal's immune function. Animal survival within the extreme environmental conditions could be in response to both artificial and natural selection that over time may leave selection signatures in the genome. The aim of this study was to identify selection signatures that may be involved in the adaptation of indigenous chickens from two different climatic regions (Brazil, a tropical state; Sri Lanka, Tropical; and Egypt, Arid) using two complimentary analyses: the fixation index (Fst) and hapFLK. Chickens from Brazil (n=156), Sri Lanka (n=92) and Egypt (n=96) were genotyped using the Affymetrix Axiom®600k Chicken Genotyping Array. Pairwise Fst analyses among countries did not detect major regions of divergence between chickens from Sri Lanka and Brazil, but detected differences between Egypt and each those populations. Moreover, common regions of difference on chromosomes 2, 3 and 8 were detected in chickens from Sri Lanka and Brazil in comparison with Egyptian chickens. The hapFLK analyses for the three separate countries suggested unique regions that are potentially under selection on chromosome 1 for all three countries, on chromosome 4 for Sri Lanka, and on chromosomes 3, 5, and 11 for the Egyptian populations. These regions contained several genes whose biological functions could provide insights in understanding adaptation mechanisms in response to arid and tropical environments.

P0512: Poultry

Comparative Analysis of Signatures of Gene Selection between Two Distinct Bird Lines using Transcriptomic Data

Modupeore Adetunji[1], Sara F. Jastrebski[1] and Carl J. Schmidt[2], (1)University of Delaware, Newark, DE, (2)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

The applications of transcriptome sequencing (RNA-seq) data in systems biology has grown exponentially over the years; such as in gene expression profiling, variant detection analysis and allele specific expression. Transcriptome data can be applied to identify functional regions influencing phenotypic traits diversity. In this study we aim to detect such signatures of selection from single nucleotide polymorphisms (SNPs) using RNA-seq data. Given the limitation of highly differential coverage between different genes, we sampled RNA-seq data from different tissues at varied time points to obtain a genome-wide scan of the entire bird, and performed comparative analysis between the different bird lines. Our results provide a template for genome-wide analysis of recent selection signatures and we found several candidate genes for recent selection similar and unique to both lines. Several of these genes are associated with growth, immunity and development. This study has direct application to future studies of animal breeding and meaningful trait selection.

P0513: Poultry

Transcriptomics Data of Liver and Adipose Tissue Highlight lncRNAs As Candidates for the Lipid Metabolism Regulation in Broilers

Kévin Franck Fabrice Muret, INRA, National Institute for Agricultural Research, Rennes, France

Broilers' body fat is not valued by food industry which must remove fat deposits on broilers' carcass during slaughter stages. Adiposity variability has a multifactorial origin with a genetic part and genes which regulate lipid metabolism are not all known yet. The aim of this study is to contribute to a better knowledge of these regulatory genes in chicken, and particularly long noncoding (lncRNA) genes, which regulate genes expression implied in a lot of diseases and metabolic processes. Therefore, we used two broilers lines (lean/fat lines) divergently selected on abdominal fat weight and fed with two different diets (high/low fat/fibers). The genotype factor leads to a differential expression (DE) of genes involved in the cholesterol synthesis and the diet factor in the hepatic fatty acid synthesis and secretion. We used RNA-seq data from two metabolic tissues: 16 livers and 16 adipose tissues and FEELnc has classified more than 6000 expressed lncRNAs. DE analyzes showed 160 (*vs.* 258) and 266 (*vs.* 396) lncRNA genes up- and down-regulated respectively between the lean and fat lines in liver (*vs.* adipose tissue). 30 (*vs.* 5) and 101 (*vs.* 1) lncRNA genes are DE between the low and the high fat diet. Amongst them, around 80 lncRNA genes are divergently localized with their nearest coding gene and potentially share a bidirectional promoter. We found some interesting examples concerning key enzymes in cholesterol and fatty acid synthesis for our both factors, genotype and diet which are good candidates for the lipid metabolism regulation in chicken.

P0514: Poultry

Spatial and Sex Differences in Gene Expression in Pectoralis Major of Broiler Chickens

Brilynn Brothers[1], Zhu Zhuo[2] and Behnam Abasht[1], (1)University of Delaware, Newark, DE, (2)Department of Animal and Food Sciences, University of Delaware, Newark, DE

Wooden Breast Disease (WBD) is a novel myopathy affecting the breast muscle of modern broiler chickens. The etiology of the disease is currently unknown, but its recent emergence has been linked to increased feed efficiency and muscle yield in broiler chickens. The cranial region of the pectoralis major tends to be more severely affected than the caudal aspect. Additionally, male chickens have higher incidence rates of WBD and tend to display more severe symptoms than females. This study aims to characterize the biological differences in the p. major between sexes of birds and regions of the muscle to determine the cause of differential susceptibility to WBD.

Samples were taken from cranial and caudal aspects of the p.major muscles of 3-week old, unaffected male and female birds for RNA sequencing. cDNA libraries were prepared, then sequenced using Illumina Hiseq2500. Sequence reads were aligned to the chicken reference genome with HISAT, then genes were analyzed for differential expression between sex and spatial groups using CuffDiff.

There were 260 differentially expressed genes between male and female birds, and 12 between cranial and caudal samples. Genes involved in fat metabolism were up-regulated in samples from the cranial region and male birds. Other significant genes included those involved in muscle

development, inflammatory processes, and oxidative stress. Results suggest increased fat deposition in male birds and the cranial aspect of the muscle. The up-regulated genes involved in oxidative stress in male birds support a hypothesis that the development of WBD is related to increased oxidative stress.

P0515: Poultry

Impact of Probiotic on Response to Heat Stress in the Broiler Chicken

Zachary S Groff[1], Matheus Reis[1] and Carl J. Schmidt[2], (1)University of Delaware, Newark, DE, (2)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

This work examined the effects that heat stress on the modern broiler chicken has on development, specifically the microbiome in the intestine along with the small intestine and spleen transcriptome. Three hypotheses were tested. First, we predicted that heat stressing will adversely affect the immune system, predominantly in the small intestine and spleen. When compared to non-heat stressed birds, we found that the data obtained from the spleen was consistent with an immune response to Leaky Gut Syndrome. Next, we hypothesized that the use of *Bacillus subtilis* will ameliorate the effects of heat stress on chickens. The data indicates that birds subjected to heat stress in the presence of *B. subtilis* had a cloacal temperature 1°C lower compared with birds subjected to the same heat stress but lacking *B. subtilis.* Finally, we hypothesized that probiotic and heat stress will impact the microbiome of the chicken's lower intestinal tract. Using 16s rRNA analysis, we examined the microbiome of the duodenum, jejunum, ileum, ceca, and large intestine. The data obtained suggests that neither heat stress nor the probiotic has any significant impact on the microbiome at the genus level.

P0516: Poultry

Comparing the Transcriptome Response to Heat Stress between Human and Avian Cell Lines

Caitlyn Hamill[1], Meaghan M. Young[1], Heidi A Van Every[1] and Carl J. Schmidt[2], (1)University of Delaware, Newark, DE, (2)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Heat stress in production animals causes major economic losses in the agriculture industry every year. The effect of heat stress on the genes of broiler chickens is not well researched. To help better understand the mechanism of heat stress in these birds we studied Human Embryonic Kidney (HEK) cells. HEK cells were grown on plates at 37°C and then random plates were selected to be heat stressed at 43.5°C.. RNAseq data was obtained from treatment and control groups. The results from this experiment will be compared to other avian cell lines from previous studies using the same heat stress protocol.

P0517: Poultry

Response of the Chicken Pituitary to Acute Heat Stress

Meaghan M. Young[1,2], Elizabeth M. Pritchett[1], Susan J. Lamont[3] and Carl J. Schmidt[4], (1)University of Delaware, Newark, DE, (2)University of Delaware, Bayville, NJ, (3)Department of Animal Science, Iowa State University, Ames, IA, (4)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Heat stress can impact gene expression patterns in multiple tissues. The pituitary gland functions as a neuroendocrine tissue modulating the body's response to stress in concert with the hypothalamus. Here we evaluate the chicken pituitary response to heat stress using transcriptome analysis. Birds were exposed to acute heat stress for two hours while control birds were maintained under thermoneutral conditions. To control for stress caused by moving birds between houses (thermoneutral to heat stress) we also moved a set of birds from one thermoneutral to another thermoneutral house. Transcriptome libraries were prepared and analyzed to identify genes that are differentially regulated by either heat stress or introduction to a new flock.

P0518: Poultry

Effect of Heat Stress on Nutrient and Amino Acid Transporters in Meat-Type Chickens

Samuel E. Aggrey[1], Walid S. Habashy[2], Marie C Milfort[1] and Romdhane Rekaya[1], (1)University of Georgia, Athens, GA, (2)Damanhour University, Egypt

Heat stress (HS) is an important stressor in poultry production and increasing global temperatures will only exacerbate its effect. We conducted an experiment to investigate the effect of heat stress on molecular transporters in broiler chickens. Forty-eight chicks at 14 d old were randomly assigned to a control group (25°C) and a HS treatment group (35°C). The Pectoralis major (*P. major*) and ileum of 5 birds per treatment were sampled at 1 and 12 d post-treatment for gene expression analysis. The total consumption and retention of amino acids, protein and fat were significantly lower in the HS group compared to the controls. In *P. major* and ileum tissues at 1 d post HS, amino acid transporters SNAT1, SNAT2, SNAT7, TAT1, and b0,+AT, were down-regulated in the HS group. The amino acid transporters B0AT and SNAT7 at d 12 post HS were down-regulated in the *P. major* and ileum. In both tissues at 1 d post HS, transporters FATP1 and SGLT1 were down-regulated in the HS group. Meanwhile, FABP1 and PepT1 were down-regulated only in the ileum of the HS group. The converse was shown in *P. major*. The nutrient transporter FABP1 at 12 d post-HS was down-regulated in both tissues, but GLUT1 and PepT2 were down-regulated only in the ileum, and PepT1 was down-regulated only in the *P. major* compared with the control group. These changes in nutrient transporters suggest that high ambient temperature might change the ileum and *P. major* lipids, glucose, amino acid and oligopeptide transporters.

P0519: Poultry

Transcriptome Analysis of Host Response to NDV Infection under Heat Stress in Two Genetically Distinct Chicken Inbred Lines

Ying Wang[1], Xianyao Li[2], Perot Saelao[1], Colin Kern[1], Kelly Chanthavixay[1], Rodrigo A. Gallardo[3], David A. Bunn[1], Terra R. Kelly[3], Susan J. Lamont[4] and Huaijun Zhou[1], (1)Animal Science, University of California, Davis, CA, (2)Shandong Agricultural University, Tai'an, China, (3)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (4)Department of Animal Science, Iowa State University, Ames, IA

Newcastle disease (ND) is an endemic and devastating disease in Africa, which can cause 80% mortality in unvaccinated village chicken flocks. Breeding chicken strains with enhanced disease resistant by identifying chicken genes or genetic markers associated with ND resistance is a promising approach to improve poultry production. The objective of the study was to identify genes and signal pathways associated with ND virus (NDV) infection under a heat stress environment in bone marrow using two highly inbred chicken lines (Leghorn and Fayoumi). At 14d of age, birds were exposed to 38°C with 50% humidity for 4 hours, then 35°C till the conclusion of the experiment at 31d of age. For the NDV treatment group, birds were inoculated at 21d with NDV La Sota strain. Total RNA was isolated from bone marrow harvested at 2 and 6 dpi and was used to identify differentially expressed genes (DEGs) with a FDR < 0.05 and fold change more than ±1.5. Preliminary analysis identified 167 and 180 DEGs at 2dpi, 139 and 153 DEGs at 6dpi in Fayoumi and Leghorn, respectively. Gene Ontology terms such as defense response to virus at 2dpi and metabolic pathways at 6dpi in Fayoumi, and negative regulation of viral genome replication at 2dpi in Leghorn birds were significantly enriched. Further bioinformatic analysis will identify more candidate genes associated with NDV infection, which can be used for potential genetic improvement of disease resistance in poultry.

P0520: Poultry

Proteomic Analysis of Liver Tissue from Heat Stressed and Newcastle Disease Virus Infected Inbred Chicken Lines.

Perot Saelao[1], Ying Wang[1], Rodrigo A. Gallardo[2], David A. Bunn[1], Terra R. Kelly[2], Susan J. Lamont[3] and Huaijun Zhou[1], (1)Animal Science, University of California, Davis, CA, (2)University of California, Davis, School of Veterinary Medicine, Population Health and Reproduction, Davis, CA, (3)Department of Animal Science, Iowa State University, Ames, IA

Newcastle disease virus (NDV) and heat stress are two major factors impacting rural poultry production in developing nations. A holistic understanding of the effects of these two combined stress factors can assist in informing and developing improved treatments through novel genetic approaches. The objective of this study was to identify specific protein and signal pathways associated with NDV infection and heat stress in two highly distinct inbred chicken lines. Two inbred chicken lines, Fayoumi and Leghorn, were treated with NDV at 21 days of age while under the constant heat exposure (35C, 65% humidity) starting from 2 weeks of age. Liver samples were collected at 2 and 6 days post-infection and flash frozen for protein analysis. Fayoumi birds had more proteins up regulated compared to Leghorn birds at both time points. Protein encoded by genes MYL4, UFC1, COLA1, and CATH1 were inversely regulated between the two lines at both time points, with Fayoumi having an increase these proteins, while Leghorn decreased its production of these proteins. Furthermore, key pathways in cell proliferation and AGE-RAGE pathways were highly differentiated between the two lines. Proteomic analysis of chicken liver samples under the effects of heat stress and Newcastle disease virus infection enables the identification of retrospective biomarkers to further investigate and characterize the response to these stressors. These results continue to highlight the unique and novel response difference between the two genetic lines and will help inform on the significant protein and response pathways in resolving NDV infection during heat stress.

P0521: Poultry

Allele-Specific Expression (ASE) of CD4+ T Cells in Response to Marek's Disease Virus Infection

Jiuzhou Song, University of Maryland, College Park, MD

Marek's disease (MD) is a T cell lymphoma disease of poultry induced by Marek's disease virus (MDV), a highly oncogenic alphaherpesvirus. To identify high-confidence candidate genes of MD genetic resistance, transcriptomic data in $CD4^+$ T cells were obtained from MDV infected and non-infected groups of two reciprocal crosses individuals mating by two highly inbred chicken lines (6_3 MD-resistant and 7_2 MD-susceptible). We identified 61 and 123 SNPs (FDR < 0.05) annotated in 39 and 132 genes in intercross $6_3 \times 7_2$ and $7_2 \times 6_3$, respectively, which exhibited allele-specific expression (ASE) in response to MDV infection. Similarly, we identified 62 and 79 SNPs annotated in 66 and 96 genes in infected and non-infected groups, respectively. Furthermore, we identified 534 and 1,543 differentially expressed genes (DEGs) (FDR < 0.05) related to MDV infection in intercrosses $6_3 \times 7_2$ and $7_2 \times 6_3$, respectively. We also identified 328 and 20 DEGs in infected and non-infected groups, respectively. After validation in $CD4^+$ T cells and tumors, we found six genes (*MCL1*, *SLC43A2*, *PDE3B*, *ADAM33*, *BLB1* and *DMB2*), especially *MCL1* gene, were highlighted as the most potential candidate genes involved in MDV infection. Many ASE genes were linked to T cell activation, T cell receptor (TCR), B cell receptor (BCR), ERK/MAPK and PI3K/AKT-mTOR signaling pathways, which played potentially important roles in MDV infection. Overall, our study also provides additional deep insights into the mechanisms of MD and disease resistance breeding in poultry.

P0522: Poultry

A Rapid-Evolving Small RNA Based Immune System Targeting Avian Leukosis Virus in Chickens

Xin Li, University of Rochester, Rochester, NY

Vertebrate genomes harbor a large number of pre-existing retroviral invaders, and face a never-ending stream of new retroviral endogenization. The challenge of controlling endogenous retroviruses (ERVs) is formidable, as ERVs can remain infectious and can recombine with other viruses or host genes to evolve into new viruses. The conflict between ERVs and their hosts is especially brutal in germ cells where the propagation of ERVs will be inherited, requiring the host to distinguish non-self from self genetic elements. The defense mechanisms essential for animal fertility from worms to humans, are PIWI-interacting RNAs (piRNAs), which protect the integrity of the germ-line genome by targeting ERVs through base-pair complementarity. However, we are just beginning to understand how the host keeps up the arms race with ever-changing ERVs. Here, we described new piRNA acquisition in vertebrates, in which chickens hijack a pre-existing provirus for piRNA production to defend against Avian leukosis virus (ALV), employing a strategy similar to the prokaryotic CRISPR-Cas system. The provirus produces anti-viral piRNAs in a domestic egg-laying breed, but does not do so in the Red Jungle Fowl, the undomesticated wild chickens. Our study identified previously unrecognized mechanisms enabling the host to rapidly evolve its piRNA repertoire and target ERVs specifically.

P0523: Poultry

Time-Series Evaluation of Wooden Breast Myopathy in Modern Broiler Chickens using RNA-Seq Analysis

Michael P. Babak[1], Erin M. Brannick[1], Carl J. Schmidt[2] and Behnam Abasht[1], (1)University of Delaware, Newark, DE, (2)Dept. of Animal & Food Sciences, University of Delaware, Newark, DE

Wooden Breast Disease (WBD), a myopathy in commercial broiler chickens that is characterized by abnormally firm consistency of the breast muscle, impacts the poultry industry negatively due to severe reduction in meat quality traits. To unravel the molecular profile associated with the onset and development of WBD in broiler chickens, we compared time-series gene expression profiles of Pectoralis (P.) major muscles between affected and unaffected birds from a high-breast-muscle-yield, purebred broiler line. This was accomplished by raising chickens for 7 weeks while harvesting P.major biopsy samples from the cranial (week 2 and 3) and caudal (week 4 and 5) aspects. Three subsets of biopsy samples comprising 4 unaffected (U) and 10 affected (A) from week 2; 4U and 12A from week 3; and 4U and 9A from week 4 were processed for RNA-sequencing analysis utilizing Illumina Hiseq platform. Sequence reads generated were processed using a suite of bioinformatics programs producing differentially expressed genes (DEGs) for each dataset. To identify genes which exhibited a more distinct expression pattern between affected and unaffected groups, DEGs were further filtered through a custom script, yielding 126, 277 and 65 genes from weeks 2, 3 and 4 respectively, at fold-change (A/U or U/A) >1.3 and FDR <0.05. Functional analysis of the filtered DEGs revealed several biological terms and pathways; top among them being metabolism, growth, inflammation, autophagy-lysosome system/ubiquitin-proteasome pathway and oxidative stress. This study, therefore, reveals presence of a unique molecular profile associated with the onset and development of WBD in commercial broiler chickens.

P0524: Poultry

Genetics of Ascites: Energy Metabolism, and Ties to Woody Breast?

Douglas D. Rhoads[1], Khaloud Alzahrani[1], Alia Parveen[1], Shatovisha Dey[1] and Nicholas B Anthony[2], (1)University of Arkansas, Cell and Molecular Biology Program, Fayetteville, AR, (2)University of Arkansas, Fayetteville, AR

Our collaborative consortium has been pursuing the underlying genetics of ascites in broilers. Recently, we used whole genome resequencing in our ascites experimental research line to identify 31 chromosomal regions with potential association with ascites phenotype. Two of those regions have now been validated for containing ascites QTLs. One region contains the CPQ gene and the other contains the LRRTM4 gene. The exact role of these genes in affecting ascites phenotype is not known. We had recently published findings that there are differences in breast mitochondrial biogenesis in breast muscle of males in the RES and SUS ascites research lines. We have now expanded this analysis to additional skeletal muscles, and to both genders. Our qPCR analyses show a marked difference in the ratio of mitochondrial to genomic sequences in breast, thigh and lung, and extreme variations between the sexes. For males, there is a significantly higher level of mitochondrial DNA in the susceptible line than in the resistant line. The difference is most dramatic in lung>thigh>breast. Others have identified differences in central metabolic protein levels when comparing normal to Woody Breast samples. Woody Breast is associated with tissue hypoxia similar to ascites. We therefore analyzed mitochondrial DNA levels in normal and Woody Breast broiler samples. Our current findings are that there is a significant difference. Surprisingly, the level of mitochondrial DNA in normal breast is more than twice that of Woody Breast (P=0.0000048). We propose that Woody Breast is a manifestation of the interaction between mild tissue hypoxia and differences in metabolic demands associated with mitochondrial biogenesis.

P0525: Poultry

Managing Bacterial Chondronecrosis with Osteomyelitis: Investigating the Mechanism of Pathogenesis in the Incidence of Lamness in Broiler Chickens

Nnamdi Simon Ekesi, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

The incidence of lameness in broiler chickens which is largely attributed to Bacterial Chondronecrosis with Osteomyelitis (BCO) is a huge burden on the poultry industry- with both animal welfare concerns, and the annual economic loss to the disease. Different studies have identified a number of reoccurring pathogens that are associated with BCO. Dr. Robert [1, 2] Wideman created the "the wire-floor model" that can induce the disease in broilers [1,2]. Using Wideman's model, we have shown that BCO can spread from bird to bird within the same pen. We start seeing lame bird on day 16, but very prominent on day 35. We have furthered this research using different parameters. We can grow birds on wire flooring, or on litter, and we can do it with and without administration of bacteria in the drinking water at day 20. On wire we get BCO lameness at 40-80% and on litter we get 20-50%. We identified a few species of Staphylococcus associated with BCO, namely S. agnetis and more recently S. hyicus species. We have been sequencing the genomes of these isolates looking for virulence genes. We tested different prebiotics and probiotics to curb lameness. We hypothesize that microbes takes advantage of compromised gut to translocate themselves into the blood and eventually make their way via the bloodstream to poorly protected growth plates of rapidly growing leg bones in broilers. We have also been looking at intestinal histopathology sections of chickens treated with different probiotics to see if there's intestinal epithelial injury or augmentation. Even though, the roles of different mucins in gut pathology after bacterial challenge is unclear, we know that Mucins are important in intestinal epithelium protection from infection. We quantified the expression of mucin and mucin-related genes. We looking at MUC 13 and MUC16 genes. We still don't understand the mechanisms that compromise gut integrity. So, we want to learn more, by treating broilers with pre and pro-biotics to see how these organisms affect gut structure and microbiome components. This study will help us understand some of the mechanisms of pathogens that compromise gut integrity-A knowledge that will lead to efficient poultry production.
Broiler; lameness; bacteria; pathogen; genome; leg.

P0526: Poultry

Direct Genome Evolution of a *Staphylococcus agnetis* Isolate, 1379, from Dairy Cattle Mammary Glands, Versus *S. Agnetis* Isolate, 908, from Bacterial Chondronecrosis with Osteomyelitis in Broilers.

Abdulkarim Shwani, University of Arkansas, Cell and Molecular Biology program, Fayetteville, AR

Abstract

Staphylococcus agnetis is a coagulase-variable, Gram positive bacterial species which has been previously associated with subclinical or mild clinical cases of mastitis in dairy cattle. This staphylococcal species has been isolated from the bone and blood of lame broilers at the University of Arkansas. Bacterial chondronecrosis with osteomyelitis (BCO) has also been successfully induced by administration of a chicken isolate of *S. agnetis* (isolate 908) in drinking water. BCO primarily affects the growth plate in the proximal femur and tibia, the fast-growing leg bones. When birds are reared on suspended wire flooring with administration of strain 908 in the drinking water (10E5 CFU/ml on days 20 and 21) lameness incidence is as high as 80% by 56 days of age. The same administration protocol induces lameness of 50% for birds raised on

litter. The annotated complete genome of strain 908 has been published. In a previous work, to better understand the relationship between dairy cattle and broiler isolates, we obtained nine *S. agnetis* isolated from milk or mammary gland secretions (n = 7) and udder skin (n = 2) from the University of Missouri for sequencing (2 x 250 MiSeq) and templated assembly (NGen ver 13, DNAStar). To trace phylogenetic relationships, we constructed phylogenetic trees based on multi locus sequence typing using either 7 housekeeping genes or 7 virulence genes. Included in this analysis were published genomes from NCBI for *S. agnetis* and the closely related species *Staphylococcus hyicus* as outgroup. The chicken isolate, strain 908, clustered with two of the cattle isolates, one of them is strain 1379 which cannot establish lameness in bird, because they failed to survive against chicken macrophages. A catalogue of gene differences between the cattle and chicken isolates is being constructed using reciprocal blast analyses at the nucleotide and polypeptide level. More than 40 genes from strain 908 are absent or poorly conserved in any of the cattle *S. agnetis* isolates. Transferring the gDNA from the strain 908infecting chicken to the strain 1379 infecting cattle, we could at least one time to transfer the pathogenic genes from the strain 908 to the strain 1379. Then the strain 1379 with the pathogenic gens of strain 908 have been subjected to macrophage experiments, in which the bacterial cells exposed to chicken macrophages to determine if the cattle isolate can get survived. After multiple trials the strain 1379 with 908 pathogenic genes could resist the macrophages, so the bacteria got a new name which is strain 1437. Work is underway on optimizing the DNA transformation to S. agnetis, and comparing the gene differences between the 1437, and each of the strains 908 and 1379.

P0527: Poultry

Genome-Wide Identification and Analysis of CTCF Binding Sites in Chickens and Pigs

Colin Kern[1], Perot Saelao[1], Ying Wang[1], Michelle M. Halstead[1], Kelly Chanthavixay[2], Ian Korf[3], Mary E. Delany[1], Hans H. Cheng[4], Juan F. Medrano[5], Alison Van Eenennaam[2], Christopher K. Tuggle[6], Catherine W. Ernst[7], Pablo J. Ross[1] and Huaijun Zhou[1], (1)Animal Science, University of California, Davis, CA, (2)University of California, Davis, Davis, CA, (3)The Genome Center, University of California, Davis, Davis, CA, (4)USDA, ARS, ADOL, East Lansing, MI, (5)University of California, Davis, CA, (6)Iowa State University, Ames, IA, (7)Department of Animal Science, Michigan State University, East Lansing, MI

Transcription factor CTCF (CCCTC-Binding factor) is a DNA binding factor that plays an important role in the regulation of 3D chromatin architecture and in defining the boundary between active and heterochromatic DNA. CTCF also plays a primary role in the activity of insulators, which block interactions between promoters and enhancers. As part of the FAANG pilot project at UC Davis, CTCF binding sites were identified across the genome using ChIP-seq in liver, lung, and spleen tissues from chickens and pigs (two biological replicates per tissue). A total of 19,699 CTCF binding sites were identified in liver, 7,658 in lung, and 12,377 in spleen in chickens. For pigs, 27,443 were identified in liver, 17,674 in lung, and 19,574 in spleen. Previous studies have found that CTCF binding sites are shared across tissues and cell types, which this study confirmed with 6,607 sites found in all three chicken tissues (86%) and 15,455 shared in pig (87%). Motif analysis of these shared binding sites revealed the CTCF motif is present in 72% of the pig binding sites, but only in 46% of chicken binding sites, indicating potential alternative motifs or CTCF behavior. *De novo* motif identification did not identify any additional motifs enriched at these locations, so further bioinformatics analysis is needed to investigate the sites that do not contain a known CTCF motif. In the future, enhancers and promoters identified from histone modification ChIP-seq data obtained for this project will be used to examine the role of CTCF as insulators.

P0528: Poultry

Identifying Host Responses to Avian Pathogenic *Escherichia coli* using the Chicken Splenic Transcriptome

Melissa S. Monson[1], Michael G. Kaiser[1], Lonneke Vervelde[2], Mark P Stevens[3], Anna Wolc[1,4] and Susan J. Lamont[1], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)The Roslin Institute, The University of Edinburgh, Midlothian, United Kingdom, (3)The Roslin Institute and Royal (Dick) School of Veterinary Studies, Easter Bush, Midlothian, United Kingdom, (4)Hy-Line International, Dallas Center, IA

Through mortality, slower growth, and decreased egg production, extra-intestinal infections with avian pathogenic *Escherichia coli* (APEC) are costly for the poultry industry worldwide. Host responses to APEC are not well understood and identifying involved genes will inform efforts to improve chicken resistance to colibacillosis. In this experiment, RNA-sequencing (RNA-seq) was used to characterize splenic transcriptomic responses to APEC in F1 birds from reciprocal crosses of broiler (disease-susceptible) and Fayoumi (disease-resistant) lines. Birds were given either APEC O1:K1:H7 or sterile PBS by intra-airsac injection. Spleen samples were collected 1 day or 2 days post infection (DPI), leading to sequencing of 48 libraries on the HiSeq 3000 (n = 6 libraries/treatment/DPI/cross). Significant differential expression between treatments and/or across DPI was detected for 440 genes, with the majority observed from exposure to APEC at 1 DPI. APEC infection increased expression of genes (such as *IL6*, *IL22*, *CCL4*, *PTX3*) predicted by pathway analysis to activate leukocytes, phagocyte differentiation and migration, and complement. By 2 DPI, many immune genes already returned to baseline expression, consistent with the decrease in bacterial load in the spleen from 1 to 2 DPI. Further mining of these datasets for allele specific expression (ASE) could provide evidence for variable *cis*-regulatory elements associated with infection with APEC. Overall, genes and pathways that respond to APEC in the spleen could contribute to innate responses, begin priming of the adaptive response, and could be targets for future investigations of resistance to APEC. Support: USDA-NIFA-AFRI US-UK Collaborative grant, Hatch project #5424.

P0529: Poultry

Genetic Analysis of Inflammatory Response Genes in the Turkey (Meleagris gallopavo)

Kadijah L. Russell[1], Jun Xu[2] and Edward Smith[2], (1)Virginia Polytechnic Institute and State University, Blacksburg, VA, (2)Virginia Tech, Blacksburg, VA

Though progress has been made in the genome analyses of the turkey, *Meleagris gallopavo*, our understanding of the genotype: phenotype relationships continue to lag behind those of other agriculturally important animal species. Among the phenotypes for which genetic understanding can be useful is inflammation, a complex trait that has been very little investigated in the turkey. Here, we report initial investigations in our lab of the genetics of inflammation in the turkey using comparative information from the chicken NOD like receptor X1 (NLRX1) and turkey Interleukin 8 (IL8). We screened IL8 and NLRX1 for nucleotide variants that may be informative for the turkey's

response to lipopolysaccharide that causes inflammation. The rationale for selecting these two genes is that IL-8 and NLRX1 have pro inflammatory and/or anti-inflammatory functions that maintain homeostasis. We designed and tested primers using three heritage turkey strains given 1.5 mg/kg LPS intra-abdominally at seven weeks of age. Differences in mortality among the three strains, involving a total of 58 birds, following LPS challenge were not significant. A total of 2,239 bp for IL8 and 572 bp for NLRX1 were screened. Six SNPs, 4 in IL8, were identified and validated. The SNPs do not appear to be correlated with response to LPS. An evaluation of the association of the SNPs with inflammation is now underway. It is hoped that further gene expression studies may provide additional insight into how variations in these and other genes are associated with differences in inflammatory response to LPS in the turkey.

P0530: Poultry

Genetic Diversity of Different Turkey Populations (*Meleagris gallovavo*)

Maria G. Strillacci[1], Erica Gorla[1], Maria C. Cozzi[1], Sergio I. Roman-Ponce[2], Felipe de Jesús Ruiz[2], Marina Duran-Aguilar[3], Vicente Murillo Vega Eliezer[4], Angel Ríos-Utrera[4], Silvia Cerolini[1], Francesca Genova[1], Maria Longeri[1] and Alessandro Bagnato[1], (1)Università degli Studi di Milano - Dept. DiMeVet, Milano, Italy, (2)Centro Nacional de Investigación en Fisiología y Mejoramiento Animal INIFAP, Ajuchitlán, Queretaro, Mexico, (3)Universidad Nacional Autonoma de Mexico, Queretaro, Mexico, (4)Campo Experimental La Posta, INIFAP, Medellín, Mexico

This study aims at investigate genetic diversity of different Italian and Mexican turkeys populations, using high-density SNP chip. The Mexico wild turkey population is considered to be the original wild type; it was brought into Europe from the Central America in the 16th century. The evolution of Mexican turkey populations, the European and the selected hybrids has been then separated by more than 500 years. A total of 116 individuals from 6 Italian breeds (Colle Euganei, Bronzato Comune Italiano, Parma e Piacenza, Brianzolo, Nero d'Italia and Ermellinato di Rovigo), 7 Narragansett turkeys, 38 commercial hybrids, 31 Mexican turkeys, were genotyped with the high-density Affymetrix 600K SNP array. Using the SNPs allele frequencie (after filtering for MAF and Hardy-Weinberg equilibrium), the genetic relationship among samples was estimated with a PCA. Genetic diversity within each population was estimated using the observed level of heterozygosity (H_o), the expected heterozygosity (H_e), the fixation index (F_{ST}), and the inbreeding coefficient (F_{IS}). Admixture analysis was also performed. After quality control, a total of 343,028 SNPs markers allowed to clearly identify the genetic structure and diversity of each turkey populations. PCA is showing a clear separation in different clusters among Italian breeds, Mexican turkeys, Narragansett and commercial hybrid. The Mexican population resulted in between the Italian populations and the commercial hybrid together with the Narragansett. Acknowledgments: Co-funded by project M01678 - Ministry of Foreign affairs of Italy and Mexico.

P0531: Poultry

Genomics Analyses of Wild × Domestic Segregation Population to Reveal Selected Genes Associated with Body Size and Plumage Color of Ducks

Zhengkui Zhou, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, China

Comparative population genomics offers an opportunity for unraveling signatures of artificial selection during farm animal domestication, however, this traditional strategy couldn't annotate their function. Based on genome-wide sequencing 1026 wild × domestic duck segregation for fine mapping, together with 106 wild and domestic genome comparison, we identified and annotated two selective sweeps with fixed new mutations harboring critical genes that are responsible for key economic traits. We found a novel intronic insertion in *MITF* accounted for white duck down feathers, and intriguingly, a long-distance regulatory mutation leading to continuously express *IGF2BP1* gene after birth can increase body size by 15% and feed efficiency by 6%. This study provides new insights to genotype-phenotype associations in animal domestication research and constitutes a promising resource for fowl economically important genes.

P0532: Poultry

Whole-Genome Resequencing Reveals Signatures of Selection and Timing of Duck Domestication

Zebin Zhang[1], Yaxiong Jia[2], Pedro Almeida[3], Judith E. Mank[3], Marcel van Tuinen[4], Qiong Wang[1] and Lujiang Qu[1], (1)Department of Animal Genetics and Breeding, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China, (2)Chinese Academy of Agricultural Sciences, Beijing, China, (3)University College London, Department of Genetics, Evolution and Environment, London, United Kingdom, (4)Centre of Evolutionary and Ecological Studies, Marine Evolution and Conservation Group, University of Groningen, Groningen, Netherlands

The genetic basis of animal domestication remains poorly understood, and systems with substantial phenotypic differences between wild and domestic populations are useful for elucidating the genetic basis of adaptation to new environments as well as the genetic basis of phenotypic change. Anas platyrhynchos (ducks or mallards) are the world's most widely distributed and agriculturally important waterfowl, and are of particular economic and importance in Asia. In order to determine the timing of duck domestication in China, as well as identify the genomic regions under selection during domestication, we performed whole genome resequencing from 78 individuals belonging to seven different duck breeds (three for meat breeds, three for egg breeds, and one dual-purpose breed) and two geographically distinct wild populations. Using the 36.1 million SNPs and 3.1 INDELs, we analyzed the structure of these populations and signatures of selection associated with domestication. We identified a complex history of domestication, with early selection for separate meat and egg lineages, originating from a single domestication event roughly 2000 years ago.Genomic comparison of wild to domesticated populations suggest that genes affecting brain and neuronal development have undergone strong positive selection during domestication. FST analysis indicates that domestication is also associated with selection for white plumage at the melanogenesis associated transcription factor locus. Our results advance the understanding of animal domestication and selection for complex phenotypic traits.

P0533: Poultry

SNP Screening and Bioinformatic Analysis of the NR3C2 Gene in Mule Duck

Shenglin Yang, Guiyang Muren Agricultural Sci-Tech Development Co., LTD, Guiyang, China

Objective:To study the possibility of *NR3C2*(Nuclear receptor subfamily 3,group C) as a candidate gene for abnormal behavior study of mule duck.Method:The PCR products of *CRHBP* gene were amplified by 9 pairs of primers of mule duck. The *NR3C2* gene and its protein structure changes was analyzed by bioinformatics software before and after SNP mutation, while the PCR product was sequenced.Results:7 SNPs were screened in the amplified *NR3C2* gene: Intron2-G13928T□Intron2-A13972G□Intron2-T14336C□Intron4-A34930G□Intron4-T35157G□ Exon8-G198088A and Exon9-G199134C. Among them, Exon8-G198088A are synonymous mutations; Exon9-G199134C are missense mutations resulting in the mutation of the encoded leucine(Leu) to valine (Val).Conclusion:The analysis of polymorphism and bioinformatics in NR3C2 gene provides theoretical founditions for further study abnormal behavior in mule duck. The results showed that the NR3C2 gene had a mutation site in the muscovy duck, which may be related to the abnormal behavior of the muscovy duck.
Key Words: mule duck; abnormal behavior; *NR3C2*; SNPs□bioinformatics analysis

P0534: Poultry

Gut Metagenomic Analysis Reveals Prominent Roles of Akkermansia and Mucispirillum in Cecal Trimethylamine Level of Duck

Xingzheng Li, Yuying Guo, Xuefeng Shi, Guiyun Xu, Ning Yang and Jiangxia Zheng, China Agricultural University, Beijing, China

The incidence of cardiovascular diseases (CVDs) is increasing globally. Recent metabolomics approaches identified that plasma trimethylamine-N-oxide (TMAO) is a novel and independent risk factor for promoting atherosclerosis (AS). TMAO generation is dependent on the gut microbiota, which first metabolizes dietary choline to trimethylamine (TMA). The aim of this study was to reveal the causative microbes, which resulted in TMA production differences in the intestinal tract. In this study, the cecal TMA level of 294 ducks after choline supplementation (4,000 mg/kg) were recorded. Ducks with contrasting cecal TMA concentration (30 birds per group) were selected to investigate their cecal microbial composition by sequencing the 16S rRNA Gene V4 region. The results showed that, overall, the microbial community in the cecum was quite similar in these two groups. However, the abundances of *Mucispirillum* were significantly higher while that of *Akkermansia* was lower (Mann-Whitney U test, $P < 0.01$; LEfSe, $P < 0.05$) in the high TMA level group. These results indicated the prominent role of cecal microbiota in TMA generation, and suggested plausible uses of *Akkermansia* to reduce TMA production. *Mucispirillum* and *Akkermansia* might act as key players in the pathogenesis of TMAO-induced CVDs. In conclusion, *Mucispirillum* and *Akkermansia* were closely related to TMA producing in the Intestinal tract.

P0535: Poultry

Characterization and Phylogenetic Analysis of Hemagglutinin Gene & Neuraminidase Genes of Avian Influenza Virus Subtype H9N2 in Pakistan

Arslan Mehboob, BIOTECHNOLOGY RESEARCH INSTITUTE CHINESE ACADEMY OF AGRICULTURAL SCIENCES 12 ZHONGUANCUN NANDAJIE HAIDIAN DISTRCIT BEIJING CHINA 100081, BEIJING, China

Despite extensive vaccination, H9 Avian influenza outbreaks has caused great economic losses to poultry industry resulting in decrease egg production, high morbidity and mortality. The ability to cross species barrier makes it a potent threat. Continuous mutations in the HA gene transforms AIV subtype H9N2 into more pathogenic virus that may have pandemic potential and can cross species barrier. Thus, it is essential to continuously monitor antigenic variants of H9 virus. HA gene plays vital role in viral attachment, release of genetic material and pathogenicity. In present study, a sum of four H9 virus samples were isolated, serological and molecular confirmation was done. 500 samples were collected and properly labelled. They were then processed for egg inoculation in embryonated eggs. Virus was grown in embryonated eggs and harvested fluid is then proceeded for confirmatory testing. Haemagllutination and Haemagllutination Inhibition testing was done. RNA was extracted by Kit method and cDNA was synthesized. Reverse Transcriptase (RT-PCR) was performed using specific primer sets and then the amplicon were run on agarose gel. PCR product was sent for sequencing and Phylogenetic tree was constructed. The present study enabled us to characterize and construct Phylogenetic tree of HA and NA gene of currently prevailing H9N2 Avian Influenza isolates in Pakistan belongs to G1-like sub lineage. Amino acid analysis revealed substitution of polybasic amino acid residues that my transform H9N2 into more pathogenic.

P0536: Poultry

AUIR: A High-Throughput Analysis Pipeline for Avian Influenza Sequence Data.

Zaid Abdo[1], Christopher Dean[1], Matthew Hopken[1] and Antoinette Piaggio[2], (1)Colorado State University, Ft. Collins, CO, (2)USDA-APHIS-National Wildlife Research Center, Ft. Collins, CO

Influenza is an infectious disease caused by RNA viruses within the Orthomyxoviridae family. These viruses cause disease in a variety of animals including birds, pigs and humans. Because of their antigenically variable nature, these viruses are able to escape the innate and adaptive immune system to cause disease -- leading to the development of seasonal vaccine strains.

With the affordability, accessibility, and high-throughput nature of next generation sequencing instruments, individual labs are now able to sequence the genomes of a variety of organisms in bulk. In the influenza domain, this has led to a number of surveillance (CDC, Nextflu, USDA), prediction (Goolge Flu Trends, Twitter Improves Influenza Forecasting), and sequence collection efforts (Influenza Research Database, Influenza Virus Database).

The purpose of AUIR is to act as a preliminary analysis platform for the analysis of large amounts of influenza sequence data. The pipeline is built with Nextflow, a parallel computational workflow language. It also uses Docker, a software containerization platform that aids in the installation of the many open-source programs utilized throughout the pipeline. Functionally, the pipeline performs various quality control checks, host-dna removal, sequence assembly, complete genome annotation, genome coverage statistics, and easy-to-view html summary reports.

P0537: Insects

Hybrid Assembly of Short and Long Reads for *de novo* Genome Assembly in Insects using MinION

Uchiyama Hironobu[1], Yu Hamaguchi[1] and Shunsuke Yajima[1,2], (1)Tokyo University of Agriculture NODAI Genome Research Center, Tokyo, Japan, (2)Department of Bioscience, Tokyo University of Agriculture, Tokyo, Japan
Insects have a large number of species. Almost all of these genomes are still unknown. Although genome size of insects varies greatly among species, these size is small relative to other animals. In general, insects have a small amount of DNA due to their body size. In addition, sequencing for a large number of insect species should be cost effective. Moreover, for high quality gene model annotation, we aim high quality genome reference (contig N50 > 10 kb, scaffold N50 > 300 kb). Considering these, we optimize de novo hybrid assembly for insects using Illumina short reads and MinION reads. We also applied Illumina synthetic long reads.
As first, a genome of Japanese honey bee, *Apis cerana japonica* were sequenced. We sequenced 3 billion of paired reads (100 bp, X100) and 250 thousand of synthetic long reads (>1500 bp, X5) by Illumina HiSeq 2500 and 500 thousand of long reads (5000 bp in average, X12) by MinION. Hybrid assembly was performed with Illumina short paired reads and synthetic long reads using SPAdes(v3.6.1). Assembled contigs were 4833 in number (Max: 910308 bp, N50: 150629 bp). For scaffolding, these contigs were combined with MinION long reads by Genome finishing module in CLC genomics workbench. Scaffolds were 1226 in number (Max: 1755329 bp, N50: 432306 bp). Genome annotation of the scaffolds with MinION long reads is better than the contigs in the score of BUSCO (v3). Now, we also challenge genome assembly in other species, a seed beetle, *Callosobruchus chinensis*.

P0538: Insects
How to Determine the Genome Sequence of Single Insects
Nikka Keivanfar[1], Preyas Shah[2], Neil Weisenfeld[2], Vijay Kumar[2], Stephen Williams[3], Claudia Catalanotti[2], Ariel Royall[1], Ben Matthews[4], Jill Herschleb[2] and David Jaffe[5], (1)10x Genomics, Inc., Pleasanton, CA, (2)10x Genomics Inc., Pleasanton, CA, (3)10x Genomics, Inc, (4)Rockefeller University, (5)10x Genomics, Pleasanton, CA
The millions of insect species are foundational to the ecology of life and as disease vectors, pollinators and pests, deeply affect humankind. Typical species have prodigious populations and are thus highly polymorphic. Insects are small, so a problem in insect genomics is extracting 'enough' DNA. One might pool individuals, but alleles would have low molarity, providing a poor substrate for genome reconstruction. Therefore, standard practice is to inbreed, an arduous and artifactually selective process. Rather than fabricating clonal strains, a better approach would reconstruct *individual* insect genomes.
We demonstrate this capability. First, we extract DNA from one insect using an optimized salting out technique, yielding molecules up to 70 kb. From this, we use ~1 ng of DNA to construct one library, capturing molecules in nanodroplets via the 10x Genomics ChromiumTM technology. The library is sequenced using short reads. Finally, SupernovaTM creates a diploid genome assembly.
To assess our method, we exploited two controls: the 1.4 Gb deeply repetitive genome of *Aedes aegypti* and the 0.2 Gb genome of *Drosophila melanogaster*. We generated assemblies from F1 hybrids having a single reference parent. These hybrids mimic polymorphic, wild type individuals. We complemented these samples with actual wild individuals captured locally. An identical turnkey laboratory and computational process was applied.
Control to reference comparison demonstrates > Q40 concordance, with scaffold sizes up to 20 Mb. We describe in detail additional assembly characteristics, including gene content. Our method can be applied routinely to the insect world to explore its biology.

P0539: Insects
A Population-Genomic Approach to Delineate Host Plant Use By an Insect Vector
Daisy (Zhen) Fu[1], Brendan Epstein[2] and William E. Snyder[1], (1)Washington State University, Pullman, WA, (2)College of Biological Sciences, Saint Paul, MN
Winged herbivorous insects can readily move between habitats and/or host plants, track high quality plant resource through space and time. Analysis of genomic variation has been one tool to understand insect dispersal patterns. The potato psyllid, *Bactericera cockerelli*, is of applied interest as a minute herbivore vectors the devastating pathogen of zebra chip disease on cultivated potato (*Solanum tuberosum*). Our previous population genomic studies (2012-13 sampling) suggested the interbreeding of psyllids hosted on potato crops and a perennial solanaceous plant, bittersweet nightshade (*Solanum dulcamara*). Meanwhile, a psyllid population from potatoes that was genetically distinct from others suggested additional potential non-crop host plants to be identified. Here we used Nextera-tagmented reductively-amplified DNA ("NextRAD") to investigate the fine-scale movement patterns of the potato psyllid from an extended sampling network in the Pacific Northwest of the USA. We identified 6,529 polymorphic loci among psyllids that were sampled between 2012 and 2016. Multiple population genetic analyses suggested that *Lycium* sp, a non-native solanaceous perennial which withstands arid environment, was more likely the source of the potato psyllids invading potato fields. This was further supported by the results that late in the season psyllids from *Lycium* sp. were more genetically similar to potato-collected psyllids than the early season comparison. Altogether, our results provide evidence that a population genomic approach form an effective means to delineate complex patterns of insect movement across landscapes.

P0540: Insects
A Highly Multiplex Amplicon Sequencing Approach for Rapid Identification of Invasive Species
Julian R. Dupuis, University of Hawaii, Hilo, HI and **Scott Geib**, USDA-ARS, Hilo, HI
Each year, thousands of exotic and invasive species are intercepted at ports of entry or detected in the environment. Rapid species identification is crucial to management and eradication of these species, and prevention of their establishment. Our goal was to develop a rapid, straightforward tool for species identification of invasive tephritid fruit flies that are commonly detected in California, Florida, and South Texas. For many species, clear morphological tools for discriminating species do not exist, particularly at the immature level. We developed a bioinformatic locus selection pipeline that takes advantage of a variety of genomic and transcriptomic data sources to identify phylogenetically-informative, conserved exons in orthologous genes. We targeted 878 conserved exons in a highly multiplexed, single tube amplicon sequencing approach for hundreds of individuals across three genera of the most economically important tephritids: *Anastrepha*, *Bactrocera*, and *Ceratitis*. This approach yielded a phylogenomic dataset that far exceeded the phylogenetic resolution of existing datasets, containing >40,000 informative characters after reasonable filtering. From this dataset, we identified the most diagnostic exons for species identification, and

evaluated the efficacy of these markers in the context of use in a diagnostic lab. The wet lab and analysis pipeline developed can analyze hundreds of individuals at a time, and return taxonomic, and in some cases population level, assignment in as few as three days from sample collection. Our approach provides a novel way to combine diverse genomic and transcriptomic data sources, particularly when at least one well-annotated data source is available, and can rapidly develop diagnostic tools for non-model systems that are scalable, cost-effective, and robust.

P0541: Insects

Genetics, Genomics, and Transcripomics of Host Specificity in Aphid Parasitoids

Keith R. Hopper, USDA-ARS, Newark, DE

Differences in parasitism among potential host species can provide strong selection for divergence and speciation in parasitic Hymenoptera. However, little is known about the evolution or genetic architecture of host specificity. Here we report research on the genetics of host specificity in *Aphelinus* species. We have sequenced, assembled, and annotated the genomes and transcriptomes of eighteen *Aphelinus* species in five complexes. Using amino acid sequences, we developed a phylogeny, onto which we mapped parasitism of seven diverse aphid species. For some aphids, parasitism was phylogenetically conserved, but for others, parasitism diverged between closely related parasitoids, consistent with host-driven speciation. To explore the genetic architecture of differences in host specificity, we crossed *Aphelinus atriplicis*, which readily parasitizes *Diuraphis noxia*, with *Aphelinus certus*, which rarely parasitizes this aphid. Using genetic markers from reduced-representation genomic libraries, we mapped quantitative trait loci (QTL) affecting parasitism of *D. noxia*. We found eight QTL that explained ~40% of the variation in parasitism among backcross females. We compared the genomes and transcriptomes of these parasitoid species to find genes that diverged in sequence or expression, and we tested whether divergent genes mapped to QTL affecting parasitism of *D. noxia*. We identified 14 divergent genes that mapped to parasitism QTL or significantly affected parasitism by themselves. One of these candidate genes is expressed in sensilla on the antennae and ovipositors of female wasps, as expected given that host specificity appears to involve recognition of aphids as suitable hosts.

P0542: Insects

A New Genome Sequence of Cotton Aphid, *Aphis gossypii*

Jongsun Park[1], Hong Xi[1], Yongsung Kim[1] and Wonhoon Lee[2], (1)InfoBoss Co., Ltd., Seoul, South Korea, (2)Gyeongsang National University, Jinju, South Korea

Aphids are one of the famous agricultural pest insects in the world. Generally, these insects damage hostplants by sucking phloem directly and transmitting plant pathogen viruses indirectly. *Aphis gossypii*, (in the tribe Aphidini and the family Aphididae), which is polypagous species, shows various morphological characters due to adaptations to different host plants, geographic isolation, and genetic drift. It will be useful to understand the differences between cotton aphid and pea aphid (*Acyrthosiphon pisum*), of which genome was already sequenced and analyzed well. Here, we generated around 180x coverage raw data of *A. gossypii* genome using Illumina HiSeq4000 with two pair-end and two mate-pair libraries. First version of draft genome presents that total length is 357.44 Mbp (N50 is 472,772 bp and max length of scaffold is 2.71 Mbp), which is similar to genome length predicted by k-mer analysis (390.11Mb). Length of most of scaffolds (310,662 out of 319,177) is less than 500bp. In comparison to pea aphid genome (*Acyrthosiphon pisum*), genome length of *A. gossypii* is two-third (541.68Mb) and GC ratio (27.73%) is similar to that of pea aphid genome (29.76%). Number of ORFs predicted by AUGUSTUS is 21,311, which is smaller than that of pea aphid (36,970 ea). Interestingly, 2,488 InterPro terms were found only in *A. gossypii*, while 744 terms were in *A. pisum*, presenting that genome composition of *A. gossypii* is quite different from *A. pisum*. All these data including genome sequences will be available in Aphid genome database (http://www.aphidgenome.info/) for further comparative genomic analyses.

P0543: Insects

RNA-Seq Analysis of *Ca.* Liberibacter asiaticus Infection of Five Developmental Stages of Asian Citrus Psyllid

Ruifeng He[1], Mark Willis[1], Tonja Fisher[2], Carol Soderlund[3], Kristen Pelz-Stelinski[4], Judith K. Brown[5] and David R. Gang[1], (1)Institute of Biological Chemistry, Washington State University, Pullman, WA, (2)School of Plant Sciences, University of Arizona, Tucson, AZ, (3)University of Arizona/Bio5 Institute, Tucson, AZ, (4)Univeristy of Florida, Citrus Research and Education Center, Lake Alfred, FL, (5)School of Plant Sciences, University of Arizona, Tucson, AZ

The pathogen bacterium *Candidatus* Liberibacter asiaticus (*C*Las) vectored by the Asian citrus psyllid (ACP) causes citrus greening disease or Huanglongbing (HLB), which is the most significant and widespread threat to the citrus industry. To investigate gene expression profiles that associate with ACP-*C*Las interactions and identify genes that respond to *C*Las infection, we constructed RNA-seq libraries from *C*Las-infected and *C*Las-free ACP samples of five different developmental stages (nymphal instars 1-2, nymphal instar 3, nymphal instars 4-5, teneral adults and post-teneral adults). Based on 1,960 million 150 bp sequencing reads totaling 296 Gbp data, we made a massive *de novo* assembly to generate 44, 667 contigs with 25,857 (57.9%) contigs being annotated, which were then further analyzed for potential functional classification and potential roles in infection. The results showed that gene expression in different developmental stages of ACP did not respond in the same manner to *C*Las infection. With more contigs being up or down- regulated, nymphal instars 4-5 and teneral adults showed a more sensitive response to *C*Las infection than nymphal instars 1-2, instar 3 and post-teneral adults. More differentially expressed genes were from the nymphal instars than adults, indicating that instars are more susceptible, e.g. responsive to defense and/or susceptibility, to *C*Las infection. This new knowledge has potential, major implications for instar-specific *C*Las abatement. In addition, specific genes with roles in development, immunity, and transmission pathways were identified. These results provide insights into ACP-*C*Las interactions and HLB control.

P0544: Insects

De Novo Assembly of the Horn Fly Transcriptome

Luisa N. Domingues[1], Lane Foil[2] and Felix Guerrero[1], (1)US Dept. of Agriculture, Agr. Research Service, Kerrville, TX, (2)Louisiana State University/Department of Entomology, Baton Rouge, LA

The horn fly, *Haematobia irritans irritans* (Linnaeus, 1758) (Diptera: Muscidae), a hematophagous ectoparasite of cattle, causes considerable economic losses to the livestock industry worldwide. This pest is mainly controlled with insecticides; however, horn fly populations from several countries have developed resistance to many of the products available for their control. The resistance mechanisms associated with pesticide resistance in horn fly include, target site insensitivity, metabolic detoxification and altered behavior. In an attempt to better understand the development of metabolic resistance in horn flies, we used an Illumina paired-end read HiSeq and GAII approach to determine the transcriptome of adult females, males, permethrin treated and permethrin + PBO treated from a Louisiana population of horn flies with a moderate level of pyrethroid resistance. A total of 128769828, 127276458, 66325598, and 63053866 quality-filtered Illumina reads were obtained for females, males, permethrin treated and permethrin + PBO treated adult flies, respectively. The *de novo* assemblies using CLC Genomics Workbench 8.0.1 yielded 17038, 12567, 2596, 7535 contigs (≥ 200 nt) for females, males, permethrin treated and permethrin + PBO treated, for which 11340, 8056, 1907 and 5835 have at least one predicted ORF, respectively. More than 95% of the ORFs of each data set had significant hits in the BlastP (Non-redundant database) (e <0.001). The number of contigs in each data set with InterPro, GO and pathway annotations were: Females - 8333, 5611, 1435; Males - 5703, 3792, 1011; Permethrin treated - 1178, 802, 225; Permethrin + PBO treated - 3514, 2355, 615.

P0545: Insects

Sequencing of an Extremophilic Ice Crawler: *Galloisiana yuasai*

Felipe A. Simão[1], Ryuichiro Machida[2], Bernhard Misof[3], Oliver Niehuis[4] and Evgeny M. Zdobnov[1], (1)University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland, (2)University of Tsukuba, Tsukuba, Japan, (3)Zoological Research Museum Alexander Koenig, Bonn, Germany, (4)Ludwig Albert University of Freiburg · Evolutionary Biology and Ecology, Breisgau, Germany

Insects are amongst the most successful and diverse animal groups, spanning over one million described species and having colonized virtually every biotop. Although in recent times the genomic sampling of insect orders has been greatly expanded, there is still a marked lack of data belonging to species that radiated in the early Triassic period, which comprise the majority of wingless insects. This sparseness of available sequencing data is due to inherent difficulties in sequencing organisms possessing both large genome sizes and high levels of heterozygosity, compounded by a lack of available inbred laboratory colonies. We took this challenge and sequenced the genome of the ice-crawler *Galloisiana yuasai* (Notoptera: Grylloblattodea: Grylloblattidae), to the best of our knowledge the first sequenced genome of this clade of wingless cryophilic insects which radiated about 250 mya. The 1.6 Gbp *G. yuasai* draft genome assembly is relatively complete, containing 95.2 % of the Arthropoda-specific BUSCO (Benchmarking Universal Single-Copy Orthologs) set of genes and currently ranks amongst the largest insect genomes sequenced so far. The genome of this wingless insect belongs to a previously unsampled insect order, Grylloblattodea. Its availability will offer new opportunities for molecular-level evolutionary and ecological studies and it will be a valuable resource for the investigation of cold-resistance mechanisms in insects.

P0546: Insects

High Quality Genome Assemblies for Pre- and Post-Harvest Pests of Cereal Crops

Erin D. Scully[1], Scott Geib[2], Nathan A. Palmer[3], Scott Sattler[3], Gautam Sarath[3] and Sheina Sim[2], (1)USDA-ARS Center for Grain and Animal Health Research, Manhattan, KS, (2)USDA-ARS, Hilo, HI, (3)USDA-ARS Wheat, Sorghum, and Forage Research Unit, Lincoln, NE

Cereal crops face pressure from insects both in the field and after harvest. While damage in the field substantially reduces yields, damage from stored product insects can reduce nutritional quality and lead to spoilage post-harvest. Despite the threats that these insects pose major to human and animal food supplies, few genomic resources are available. However, the inability to obtain sufficient concentrations of high-quality DNA has impeded the rapid development of genome assemblies for many economically devastating agricultural pests. Using 10X Chromium libraries, we obtained high quality draft assemblies of three major pre-harvest aphid pests of cereal crops (*Sipha flava* (yellow sugar cane aphid), *Melanaphis sacchari* (sugar cane aphid), and *Schizaphis graminum* (greenbug)) and their microbial symbionts with total assembly lengths that exceeded 90% of the estimated genome sizes. In addition, we have also used these approaches to sequence the genomes of stored product pests belonging to six different families. In these cases, high molecular weight DNA isolated from individual insects was used to generate draft assemblies with scaffold N50s ranging from 500 kb to 1.7 Mb, max scaffold lengths ranging from 1.5 Mb to over 10 Mb, and recovery of over 93% complete/single copy BUSCOs. High density linkage maps and MinION long-read sequencing are being incorporated into these assemblies for superscaffolding and gap filling, respectively. These genomes will greatly expand our understanding of the physiological capacities of these insects and will lead to the identification of genetic factors that contribute to host range and tolerance to biotic and abiotic stresses.

P0547: Insects

Transcriptome Response of Leafhopper Plant Virus Vectors at Different Temperatures

Junhuan Xu[1], Lucy R. Stewart[2], Jane Todd[2], Margaret G. Redinbaugh[2] and Kwang-Ho Kim[3], (1)Plant Pathology, 023 Selby Hall, The Ohio State University, Wooster, OH, (2)USDA-ARS Corn, Soybean and Wheat Quality Research Unit, Wooster, OH, (3)National Institute of Agricultural Sciences, Crop Protection Division 166, Wanju-gun, Jeollabuk-do, South Korea

The majority of plant pathogenic viruses are transmitted to their host plants by insect vectors. The black-faced leafhopper (*Graminella nigrifrons*) and the corn leafhopper (*Dalbulus maidis*) are two insects that transmit maize rayado fino virus (MRFV). MRFV infects both maize and its leafhopper vectors, which transmit the virus in a persistent-propagative manner. Transcriptomes and expression response to exposure to other viruses are available for *G. nigrifrons*, but very little sequence data are currently available for *D. maidis*. However, the molecular mechanisms and impact of environmental factors such as temperature associated with virus transmission are not well defined. The effects of temperature (25°C and 30°C) and time (4 h and 7 days) on virus transmission and vector gene expression were examined after moving leafhoppers to healthy maize and sampling groups of each leafhopper species under the four different conditions, respectively. Four replicates for each temperature/time pair were carried out. In total 64 samples (160 leafhoppers each) were collected, the Illumina sequenced mRNA to

generate 279 GB of sequence data with a total read count ca. 3,900,000,000 with average read length of 100 bp. From these data, we generated a transcriptome assembly for *D. maidis*, and began to compare virus responsive genes of the two leafhoppers.

P0548: Insects

Flower-Foraging Insects and their Pollen Loads in French Permanent Grasslands

Patricia Faivre Rampant[1], Anne Farrugia[2], Jean Noel Galliot[3], Alice Michelot-Antalik[4], Sandra Novak[5], Aurélie Chauveau[1], Aurélie Bérard[1], Isabelle Le Clainche[1], Jean François Odoux[6], Jean Louis Fiorelli[7], Laurent Lanore[2], Nadia Michel[4], Marie-Christine Le Paslier[1] and Dominique Brunel[1], (1)INRA, US1279 Etude du Polymorphisme des Génomes Végétaux, F-91000, Evry, France, (2)INRA, UMRH 1213, Centre Auvergne - Rhône-Alpes, F-63122, Saint-Genès-Champanelle, France, (3)INRA, UREP 0874, Centre Auvergne - Rhône-Alpes, F-63122, Clermont Ferrand, France, (4)INRA, Université de Lorraine, UMR 1121 LAE Agronomie et Environnement, F-54500, Vandoeuvre les Nancy, France, (5)INRA, FERLUS 1373, Centre de recherche Nouvelle-Aquitaine-Poitiers, F-86600, Lusignan, France, (6)INRA, Entomologie 1255, Centre de recherche Nouvelle-Aquitaine-Poitiers, F-86600, Surgères, France, (7)INRA, UR SAD, Centre Grand Est, F-88000, Mirecourt, France

Semi natural grasslands are considered as a vital habitat for wild pollinators, which in return contribute to preserve the floristic diversity of this environment. To study the interactions between pollinators and plants, flower-foraging insects were caught from beginning of May to end of July along three contrasted dairy farming systems in France. Sampling was carried out along six walking transects for each farming system. We developed and test in parallel a method based on DNA barcoding analysis, allowing a quick identification of the insect and its pollen load at the same time. The results from more than 1000 flower visitor insects support the idea that DNA metabarcoding provides accurate information about the plants-insects networks. We showed spatial and temporal variation between the 3 systems and also between the 18 grasslands. DNA barcoding that most of the collected insect carried more than two plant genera. It also pointed out sensitive issues, especially the necessity to build reliable international barcode databases.

P0549: Brassicas, Arabidopsis, and related

Improving the Management of Brassica Genetic Resources in the BrACySol Biological Resource Center

Florence Esnault[1], Alain Label[1], **Muriel Thomasset**[2], Sylvain Théréné[1], Stéphane Doré[1], Philippe Duffé[3], Valérie Roussel[4], Anne Bernole[2], Arnaud Remay[2] and Anne-Marie Chèvre[3], (1)IGEPP, Institut National de la Recherche Agronomique (INRA), Agrocampus Ouest, Université Rennes 1, F-29260, Ploudaniel, France, (2)GEVES - BioGEVES, F-17700, Surgeres, France, (3)IGEPP, Institut National de la Recherche Agronomique (INRA), Agrocampus Ouest, Université Rennes 1, F-35650, Le Rheu, France, (4)Université Bretagne Occidentale, F-29200, Brest, France

INRA BrACySol Biological Resource Center (BRC) conserves a large *Brassica* collection which includes around 900 *B. oleracea* populations and 3600 rapeseed lines. Cabbage populations have been collected in farms before the massive use of hybrids. These accessions are conserved as frozen seeds. Seed batches are renewed every 10 – 15 years according to their germination rate. Project 'SecureBracysol', funded by GIS IBiSA, aimed at improving the management of this collection by dealing with two issues.

On the one hand, in order to identify putative duplicates, a molecular fingerprinting of all accessions was performed. For each population, two bulks of 20 seeds have been genotyped by BioGEVES with 9 SSR markers chosen for their allelic diversity and genome coverage. Genetic distances using Dice dissimilarity index have been calculated to compare the two bulks of each population and to compare populations between each other. The two bulks gave similar profiles for 55% of the populations, indicating an inter-bulk diversity. Some populations that could not be distinguished need to be further analysed with additional SSR markers.

On the other hand, genetic drift from one multiplication cycle was evaluated by analysing 9 accessions chosen to represent 3 cultigroups (cauliflower, cabbage, kale). For each accession, 90 plants from each of two generations were genotyped by BioGEVES with 17 SSR chosen using same criteria as above. Several genetic diversity indices were calculated and showed that two generations did not differ significantly for each analysed accession. These results validate the multiplication process which is in use.

P0550: Brassicas, Arabidopsis, and related

Landscape of Gene Transposition-Duplication Events in the Brassicaceae Family and their Implications in Adaptation

Dong-Ha Oh and Maheshi Dassanayake, Louisiana State University, Baton Rouge, LA

Gene duplication, deletion, and transposition erode co-linearity of orthologs between genomes over time, providing a substantial source of genome structural variation. We developed a pipeline to systematically trace all gene transposition and duplication events that are unique to a lineage or shared by lineages, by building Ortholog Networks (OrthNets) based on co-linearity among orthologs from multiple closely related genomes. Using this pipeline, we explored evolutionary events that result in erosion of co-linearity among six genomes in Brassicaceae (crucifer or cabbage family), including the model plant Arabidopsis and two extremophyte wild-relatives of Brassica crops adapted to highly saline habitats. Out of 17,432 OrthNets derived from the six Brassicaceae genomes, 7,034 consisted of six co-linear single-copy orthologs from each genome, while the remaining included various degrees and combinations of gene duplication, deletion, and transposition. Here, we focused on transposition-duplication (*tr-d*) events that provided a mechanism for variations in both gene copy number and co-linearity. We identified subsets of lineage-specific *tr-d* events with signatures of selective retention and sub-functionalization in all six genomes. These included LS *tr-d* of genes that may be critical for the local adaptation of extremophytes, such as orthologs of *SALT TOLERANCE 32* and *ZINC TRANSPORTER 3*. Among *tr-d* events uniquely shared by the two extremophytes, those resulted in duplication of putatively functional full-length ORFs were mostly originated from their common ancestor rather than parallel independent events. Our approach enabled systematic identification of a complex mosaic of evolutionary events among multiple close-related genomes, as well as candidate genes that may have played roles in the adaptation uniquely found in specific lineages.

P0551: Brassicas, Arabidopsis, and related

Genome-Wide Analysis of NBS-LRR Genes in the Brassicaceae and Applications for Breeding

Fabian C. Inturrisi[1], Soodeh Tirnaz[1], Philipp E. Bayer[1], Ting Xiang Neik[1], Hua Yang[1,2], Aria Dolatabadian[1], Fangning Zhang[1], Anita Severn-Ellis[1], Dhwani A. Patel[1], Aneeta Pradhan[1], HueyTyng Lee[1], David Edwards[1] and Jacqueline Batley[1], (1)University of Western Australia, Perth, Australia, (2)University of Queensland, Brisbane, Australia

The Brassicaceae family contains some of the world's most important economic and agronomic crops, which are utilised as edible and industrial oilseeds (e.g *Brassica napus, B. juncea*) and vegetables (e.g *Brassica oleracea, Raphanus raphanistrum*), along with the scientific model plant *Arabidopsis thaliana* and highly diverse wild species. Pathogens, such as *Leptosphaeria maculans* (causal agent of Blackleg) and *Sclerotinia sclerotiorum* (causal agent of Sclerotinia stem rot), severely affect production of important crop species from the Brassicaceae. Plant genomes harbour resistance (R) genes, which play an important role in plant immunity, where nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes are the most common type of R gene. In this study, NBS-LRR genes were identified and classified in 35 wild and cultivated Brassicaceae species to determine the phylogenetic relationship and better understand the long-term evolutionary history of these R genes. The variation of classes of NBS-LRR genes was observed to vary greatly within and between species, and among different genome assemblies of the same species. The expansion and loss of NSB-LRR genes was also observed among the Brassicaceae. The change in climate will have an, as yet, unknown impact on pathogen populations, with the potential for disease pressure to increase. This analysis provides a valuable resource for the identification of R genes for enhanced crop protection by developing elite resistant cultivars where by the functionality of R genes is studied against particular diseases and bred into commercial cultivar by marker-assisted breeding.

P0552: Brassicas, Arabidopsis, and related

Scaffolding Brassinosteroid Signaling Components at the Plasma Membrane by TTL Proteins

Vitor Amorin-Silva[1], Alvaro Garcia-Moreno[2], Araceli Castillo[1], Naoufal Lakhssassi[3], Jessica Perez-Sancho[1], Yansha Li[4], Alicia Esteban[1], David Pose[5], Josefa Perez-Rodriguez[6], Jinxing Lin[7], Victoriano Valpuesta[8], Omar Borsani[9], Cyril Zipfel[10], Alberto Macho[4] and **Miguel A. Botella**[8], (1)Universidad de Malaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Malaga, Spain, (2)Universidad de Malaga-Consejo Superior de Investigaciones Científicas, Malaga, Spain, (3)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (4)Shanghai Center for Plant Stress Biology, Shanghai, China, (5)Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, Malaga, Spain, (6)1Departamento de Biología Molecular y Bioquímica, Malaga, Spain, (7)5College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, (8)IHSM-University of Málaga-CSIC, Málaga, Spain, (9)Departamento de Biología Vegetal, Laboratorio de Bioquímica, Facultad de Agronomía, Montevideo, Uruguay, (10)The Sainsbury laboratory, Norwich, United Kingdom

Brassinosteroids (BRs) form a group of steroidal hormones essential for plant growth, development and stress responses. BR perception at the plasma membrane initiates a series of phosphorylation events enabling the nuclear accumulation and activity of the key transcription factors BZR1 and BES1. We found that plant-specific Tetratricopeptide Thioredoxin-Like (TTL) proteins are positive regulators of BR signaling that function as scaffold for the BR signaling components in Arabidopsis. TTL3 associates with most core components involved in BR signaling, with the exception of the BAK1 co-receptor. TTL3 is mainly localized in the cytoplasm, and BR treatment increases its localization at the plasma membrane. In addition, the expression of *TTL3* strengthens the association of BR-signaling components BSK1 and BZR1 at the plasma membrane. Consistent with a role in BR signaling, mutations in *TTL3*, and related *TTL1* and *TTL4* genes cause reduced BR responses, and these defects that highly enhanced in a triple *ttl1 ttl3 ttl4* mutant. We propose a novel mechanistic model for BR signaling, in which cytoplasmic/nuclear BR components bound to TTL proteins are recruited to the plasma membrane upon BR perception, which in turn allows the assembly of a BR signaling complex with the goal of ensuring de-phosphorylation and nuclear accumulation of the transcription factors BZR1 and BES1. This novel TTL scaffold model for BR signaling resembles that of Wnt signaling in metazoans, in which TTL proteins would act similar to Axin1, optimizing signaling efficiency of the cascade by promoting the assembly of the signaling complex at the plasma membrane.

P0553: Brassicas, Arabidopsis, and related

Chemical Genetics Dissection of Interference between Pathogen and Drought Stress Tolerance Signaling in Plants

Jiyoung Park[1], Tae-Houn Kim[1,2], Aaron Stephan[1], Rebecca Schwab[3], Detlef Weigel[3] and Julian I. Schroeder[1], (1)University of California, San Diego, La Jolla, CA, (2)Duksung Women's University, Seoul, South Korea, (3)Max Planck Institute for Developmental Biology, Tübingen, Germany

The plant hormone abscisic acid regulates adaptation to environmental stresses, particularly drought. How plants cope with multiple stresses, especially when challenged with pathogen infection and then drought, remains largely unknown. The tolerance mechanisms against the two stresses often negatively affect each other. However, the underlying mechanisms remain unknown. Using a chemical genetics approach that can address genetic redundancy and network robustness, a novel small molecule "DFPM" was identified that down-regulates abscisic acid signaling by activating plant immune responses (1-3). To dissect this interference signaling, an Arabidopsis thaliana reporter line harboring an ABA-inducible marker pRAB18:GFP was EMS mutagenized and screened for hyposensitive responses to DFPM. *rda* (***r***esistant to ***D***FPM inhibition of ***A***BA signaling) mutants were isolated and mapped to a putative receptor-like kinase. Further characterization of the functions of this receptor-like kinase in plant immune signaling and interference mechanisms with ABA signaling will be presented. This research will help understand how plants exposed to both pathogen and drought can coordinate effective tolerance responses which will be relevant for plant survival and crop yield.

1. H. Kim et al., Chemical genetics reveals negative regulation of abscisic acid signaling by a plant immune response pathway. ***Current Biology*** 21:990-997 (2011

2. H. Kim et al., Natural Variation in Small Molecule–Induced TIR-NB-LRR Signaling Induces Root Growth Arrest via EDS1- and PAD4-Complexed R Protein VICTR in Arabidopsis. ***The Plant Cell*** 24: 5177-5192 (2012

3. H. Kunz, J. Park et al., Small Molecule DFPM Derivative-Activated Plant Resistance Protein Signaling in Roots Is Unaffected by EDS1 Subcellular Targeting Signal and Chemical Genetic Isolation of victr R-Protein Mutants. ***PLoS One*** 11: e0155937 (2016)

P0555: Brassicas, Arabidopsis, and related

Arabidopsis tRNA Atlas

Maciej Szymanski[1], Patrycja Plewka[2], Przemyslaw Nuc[2], Artur Jarmolowski[2] and **Wojciech M. Karlowski**[1], (1)Department of Computational Biology, Adam Mickiewicz University, Poznan, Poland, (2)Department of Gene Expression, Adam Mickiewicz University, Poznan, Poland

Annotation is one of the most important and tedious task in genomic research. This step however, defines the success of majority of subsequent applications of the genomic data. Among functional genetic elements, the non-coding RNAs are still the most difficult to properly annotate, especially when only computational methods are available. tRNA is a primary example among such non-coding RNAs of which annotation is based often solely on computer predictions. tRNAs are the oldest evolutionary molecules that play pivotal role in process of translating information encoded in mRNA into protein. Recent studies, however, revealed novel functions for these molecules including signalling, transcription initiation and generation of short RNAs - so called tRNA-derived sRNA fragments (tRFs). In most cases the annotation strategy (e.g. tRNAScan-SE) is based on well described structural properties of the secondary clover-leaf tRNA structure. However, despite being very effective such approach will lead to identification of molecules that closely resemble well-defined tRNA model. They will fail to annotate tRNAs with unusual properties and characteristics (so called tRNA-lookalikes). Here, we report a combined experimental and computational strategy to annotate the full complement of tRNA-space of model plant *Arabidopsis thaliana*. By applying specially designed tRNA-seq procedure to identify functional molecules we have evaluated the current annotation of tRNA genes. Additionally, we were able to identify novel tRNA-like sequences that possess some of the tRNA characteristics and seem to participate in tRNA-related pathways. The used methodology and current status of annotation of the *Arabidopsis thaliana* tRNA-space will be presented.

P0556: Brassicas, Arabidopsis, and related

The AraGWAS Catalog: A Curated and Standardized *Arabidopsis thaliana* GWAS Catalog

Arthur Korte, Center for Computational and Theoretical Biology, University Wuerzburg, Wuerzburg, Germany

The abundance of high-quality genotype and phenotype data for the model organism *Arabidopsis thaliana* enables scientists to study the genetic architecture of many complex traits at an unprecedented level of detail using genome-wide association studies (GWAS). GWAS have been a great success in A. Thaliana and many SNP-trait associations have been published. With the AraGWAS Catalog (https://aragwas.1001genomes.org) we provide a publicly available, manually curated and standardized GWAS catalog for all publicly available phenotypes from the central A. Thaliana phenotype repository, AraPheno (https://arapheno.1001genomes.org). All GWAS have been recomputed on the latest imputed genotype release of the 1001 Genomes Consortium using a standardized GWAS pipeline to ensure comparability between results. The catalog includes currently 167 phenotypes and thousands of significant SNP-trait associations.

P0557: Brassicas, Arabidopsis, and related

Characterization of AtAPOSTART Double Mutants

Domenico Aiello[1], Gianpiero Marconi[2] and Emidio Albertini[2], (1)University of Perugia, Perugia, ID, Italy, (2)University of Perugia, Perugia, Italy

Apomixis is a naturally occurring mode of asexual reproduction in flowering plants, that allows the inheritance and perpetuation of the maternal genome through seed by circumventing genome re-assortment due to meiosis and fertilization. In Poa pratensis we isolated a gene, named APOSTART, putatively involved in programmed cell death. Therefore, it could be involved in the non-functional megaspore and nucellar cell degeneration that permit the enlargement of maturing embryo sacs. To better understand APOSTART function and its putative role in apomixes, we are characterizing it in Arabidopsis thaliana. PpAPO shares high homology with two Arabidopsis proteins: AtAPOSTART1 (AtAPO1), and AtAPO2. In order to verify if AtAPO1 and AtAPO2 have additive or redundant roles we generated and analyzed the atapo1/atapo 2 double mutants. Double mutants germinated slower and plants appeared smaller than wt. Moreover, high variability was found between and within mutant lines.

The reason for this variability could be found in polyploidy or epigenetic changes. With this aim on mind, we performed a comparison of DNA methylation between double mutants and wt for AtAPO promoter regions as well as for their reduced representation whole genomes. On the other hand, polyploidy was investigated both by checking the chromosome number of some genotypes and by flow cytometry. Overall results in terms of the methylation analysis and ploidy in the Arabidopsis double mutants are reported and critically discussed.

P0558: Brassicas, Arabidopsis, and related

Extensive Transcriptomic and Epigenomic Remodeling occurs during Arabidopsis Germination

Reena Narsai[1], Quentin Gouil[2], David Secco[3], Akanksha Srivastava[3], Yuliya Karpievitch[3], Lim Chee Liew[1], Ryan Lister[3], James Whelan[1] and **Mathew G. Lewsey**[1], (1)La Trobe University, Bundoora, Australia, (2)Walter and Eliza Hall Institute, Melbourne, Australia, (3)University of Western Australia, Perth, Australia

Seed germination involves progression from complete metabolic dormancy to a highly active, growing seedling. Many factors regulate germination and these interact extensively, forming a complex network of inputs that control the seed-to-seedling transition. Our understanding of the direct regulation of gene expression and the dynamic changes in the epigenome and small RNAs during germination is limited. The interactions between genome, transcriptome and epigenome must be revealed in order to identify the regulatory mechanisms that control seed germination.

We present an integrated analysis of high-resolution RNA sequencing, small RNA sequencing and MethylC sequencing over ten developmental time points in Arabidopsis thaliana seeds, finding extensive transcriptomic and epigenomic transformations associated with seed germination. We identify previously unannotated loci from which messenger RNAs are expressed transiently during germination and find widespread alternative splicing and divergent isoform abundance of genes involved in RNA processing and splicing. We generate the first dynamic transcription factor network model of germination, identifying known and novel regulatory factors. Expression of both microRNA and short interfering RNA loci changes significantly during germination, particularly between the seed and the post-germinative seedling. These are

associated with changes in gene expression and large-scale demethylation observed towards the end of germination, as the epigenome transitions from an embryo-like to a vegetative seedling state.
This study reveals the complex dynamics and interactions of the transcriptome and epigenome during seed germination, including the extensive remodelling of the seed DNA methylome from an embryo-like to vegetative-like state during the seed-to-seedling transition.

P0559: Brassicas, Arabidopsis, and related

A Serine-Rich Duplicated Region of ASY3 Confers Meiotic Stability in Autotetraploid *Arabidopsis lyrata*

Paul J. Seear[1], Sarah Marburger[2], Roswitha Schmickl[3], Kirsten Bomblies[2], Levi Yant[2] and James D. Higgins[1], (1)University of Leicester, Leicester, United Kingdom, (2)John Innes Centre, Norwich, United Kingdom, (3)Czech Academy of Sciences, Pruhonice, Czech Republic

Whole genome duplication (WGD) is often associated with increased ecological fitness and adaptation to new biological niches. However, the doubled set of chromosomes can lead to complex meiotic configurations at meiotic metaphase I, thus causing sterility. Our aim is to understand the mechanisms underlying the stabilization of bivalent formation during meiosis after WGD in naturally occurring stable polyploid plants. *Arabidopsis lyrata* is an outbreeding relative of the model plant *A. thaliana* that has both extant diploid and stable autotetraploid populations. Whole genome resequencing of diploid and tetraploid populations of *A. lyrata* has showed strong evidence of selection on a number of synaptonemal complex genes. Sanger sequencing of the coding region of these genes revealed two major ASY3 (Red1 homolog) alleles in diploids and tetraploids. One has an in-frame duplication of a 26aa serine rich region (TD allele), the other has a reference genome-like ASY3 allele (ND allele). PCR genotyping of these alleles and cytological analysis in a number of diploid and tetraploid populations showed that tetraploids homozygous for the TD allele were meiotically stable (bivalents with distal chiasmata), whereas tetraploids with both TD and ND alleles formed multivalents and univalents. Diploid *A. lyrata* populations were almost always homozygous for the ND allele and displayed stable bivalents with a large proportion of proximal/interstitial chiasmata. Computer based predictions of both *in silico* translated ASY3 alleles indicated Ataxia telangiectasia mutated (ATM) sites in the duplicated region, with the TD allele harboring potentially double the sites than the ND allele in that region.

P0560: Brassicas, Arabidopsis, and related

WRKY7, -11 and -17 Transcription Factors, are Repressors of Unfolded Protein Response Genes through bZIP28 Transcription Factor, during PAMP-Triggered Immunity in *Arabidopsis thaliana*

Adrian Moreno[1], Ariel Herrera-Vasquez[2], Jose Dominguez[3], Jesus Vicente-Carbajosa[3], Joaquin Medina[3], Claudio Meneses[4] and **Francisca Blanco**[5], (1)Facultad de Ciencias Biologicas, Universidad Andres Bello, Santiago, Chile, (2)Facultad de Ciencias Biologicas, Pontificia Universidad Catolica de Chile, Santiago, Chile, (3)Centro de Biotecnologia y Genómica de Plantas Universidad Politécnica de Madrid, Madrid, Spain, (4)FONDAP, Center for Genome Regulation, Santiago, Chile, (5)Centro de Biotecnología Vegetal, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago 837-0146, Chile, Santiago, Chile

Plants have evolved sophisticated mechanisms to protect themselves from pathogens. The recognition of pathogens triggers a signaling cascade, which leads to the biosynthesis of salicylic acid, ROS, callose and the upregulation of pathogenesis-related proteins (PRs). The accumulation of PRs generates endoplasmic reticulum stress and triggers the unfolded protein response (UPR). However, failure to attenuate the UPR may have detrimental effects on plants. In this context, WRKY7, 11 and 17 transcription factors play a key role in the regulation of the defense response since mutant plants lacking these factors are more resistant to pathogenic bacteria infection.To get insights about the molecular mechanisms involved in WRKYs mutant resistance phenotype and UPR, we analyzed the expression of the main components of the signaling pathways suggesting that bZIP28 transcription factor plays a key role in the regulation of ER chaperones upon plant exposure to Flg22. We show that triple *wrky*-mutant plants are more effective at establishing a defense response against *Pst* DC3000 infections. Additionally, triple *wrky* mutant plants exhibited a larger number of callose deposits and accumulates more transcript of ER chaperones in response to Flg22. Also, *wrky* mutant accumulates more transcript of *bZIP28* suggesting that WRKY7, 11 and 17 acts as transcriptional repressors of this gene. Using Arabidopsis protoplast assays we tested if WRKY7, 11 and 17 can bind bZIP28 promoter trough W box *cis*-elements. Based on these results, we postulate a fine-tuning model of basal defense response regulation in Arabidopsis, including the negative control of gene expression associated with UPR genes controlling the physiological response of plant-pathogen interaction.
Grant Fondecyt Regular N^{o} 1170259.

P0561: Brassicas, Arabidopsis, and related

Comparison of *in vivo* and *in vitro* Targets of *AtbZIP28*

Taiji Kawakatsu, National Agriculture and Food Research Organization, Tsukuba, Japan

Unfolded protein accumulating in the lumen of the endoplasmic reticulum (ER) triggers ER stress and activates ER-to-nucleus signaling pathways to alleviate the stress. In *Arabidopsis*, three bZIP transcription factors, AtbZIP17, AtbZIP28 and AtbZIP60 are involved in this process, but little is known about gene regulatory network in ER stress response in Arabidopsis. Here, I compared the direct targets of ER membrane associated AtbZIP28 identified by *in vivo* ChIP-seq and *in vitro* DAP-seq using publicly available sequencing data. 133 genes and 1,469 genes were identified as the direct targets of AtbZIP28 *in vivo* and *in vitro*, respectively. We will present the genomic and epigenomic signatures of consistent and inconsistent targets between *in vivo* and in *vitro*.

P0562: Brassicas, Arabidopsis, and related

Arabidopsis Shoot Meristem-Specific Analysis of Gene Expression and Histone Modification Patterns and their Temporal Dynamics during Flowering

Aneta Sawikowska[1,2], Yuan You[3,4], Manuela Neumann[4], David Pose[5], Giovanna Capovilla[4], Tobias Langenecker[4], Richard A. Neher[6], Markus Schmid[4,7] and **Paweł Krajewski**[2], (1)Department of Statistical and Mathematical Methods, Poznań University of Life Sciences, Poznan, Poland, (2)Department of Biometry and Bioinformatics, Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland, (3)Zentrum für Molekularbiologie der Pflanzen, Eberhard-Karls-Universität Tübingen, Tübingen,

Germany, (4)Max Planck Institute for Developmental Biology, Department of Molecular Biology, Tübingen, Germany, (5)Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, Malaga, Spain, (6)Biozentrum, University of Basel, Basel, Switzerland, (7)Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, Umeå, Sweden
While the completed genomic sequences of many organisms foster gene discovery, our knowledge of gene functions and regulation is increasingly lagging. Like other eukaryotes, plants are composed of many specific cell types that make function-distinct tissues. How gene expression is precisely controlled and how epigenomic modifications add on the genomic information to regulate transcription locally has eluded analysis so far, mostly because of the poor accessibility of many plant tissues. For example the shoot apical meristem (SAM) that possesses a number of pluripotent stem cells located at the growing tip can produce organs throughout the entire life of plants. At the time of flowering, the SAM of *Arabidopsis thaliana* switches fate and starts producing flowers instead of leaves. Correct timing of flowering in part determines reproductive success, and is therefore under environmental and endogenous control. Here we report the temporal dynamics of the chromatin modifications H3K4me3 and H3K27me3 and their correlation with transcriptional changes at the SAM in response to photoperiod-induced flowering. Emphasizing the importance of tissue-specific epigenomic analyses we detect enrichments of chromatin states in the SAM that were not apparent in whole seedlings. Furthermore, our results suggest that regulation of translation might be involved in adjusting meristem function during the induction of flowering. In this presentation we also comment on the properties of computational tools that are necessary to uncover the meaningful biological interpretation of obtained high-throughput sequencing data.
You et al. (2017). Temporal dynamics of gene expression and histone marks at the *Arabidopsis* shoot meristem during flowering. Nature Communications 8: 15120, doi:10.1038/ncomms15120.
Research supported through the ERA-CAPS project 'FlowPlast'

P0563: Brassicas, Arabidopsis, and related
Regulation of the Embryo-to-Seedling Phase Transition By a Scarecrow-like Protein
Ming-Jun Gao[1], Xiang Li[2], Jun Huang[2], Gordon M. Gropp[2], Qi Chen[3], Z. Jeffrey Chen[4], Abdelali Hannoufa[5] and Dwayne D. Hegedus[2], (1)Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, Saskatoon, SK, Canada, (2)1. Agriculture and Agri-Food Canada, Saskatoon Research and Development Centre, Saskatoon, SK, Canada, (3)2. State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China, (4)The University of Texas at Austin, Austin, TX, (5)Agriculture and Agri-Food Canada, London Research and Development Centre, London, ON, Canada
The initiation and termination of the seed maturation phase and transition to germination and vegetative growth depend on the maintenance of cell fates and the correct deployment of developmental programmes. Many of the developmental processes active during seed maturation and seed filling are repressed after seed germination and seedling establishment. Epigenetic regulation has been implicated in repressing embryonic traits during seed germination and vegetative growth. HISTONE DEACETYLASE19 (HDA19) was shown to form a multi-protein complex for the repression of gene expression and to be involved in the repression of embryonic properties after germination, although the underlying mechanism remains unclear. We identified SCARECROW (SCR)-LIKE15 (SCL15) as being physically associated with HDA19 and required for the repression of a large subset of seed maturation genes. We provide evidence that ectopic expression of embryonic genes in *scl15-1* seedlings correlates with the histone H3 hyper-acetylation of chromatin at seed-specific loci; some of these loci are identified as direct targets of HDA19–SCL15 association. Mutation of *SCL15* affects seed vigor and germination and increase sensitivity to abscisic acid during germination and post germination growth. Moreover, SCL15 was identified as a positive regulator of primary seed dormancy in *Arabidopsis*. These findings suggest that SCL15 acts as part of an HDA19-associated complex to regulate seed dormancy, germination and seed-to-seedling phase transition.

P0564: Brassicas, Arabidopsis, and related
Small RNAs from Diverse Sources Target Reactivated Transposable Elements in *Arabidopsis thaliana*
Kaushik K. Panda, The Ohio State University, Columbus, OH
Transposable elements (TEs) are mobile genetic units found in all eukaryotic genomes. TE insertions and/or rearrangements cause mutations and DNA damage. To defend their genome, organisms have evolved various silencing mechanisms to repress TE activity. Small RNA-directed DNA methylation (RdDM) is one such silencing mechanism that is well-studied in the reference plant Arabidopsis. Canonically, RdDM functions through RNA Polymerase IV (Pol IV) generation of 24 nucleotide small RNAs (24nt sRNAs) that are incorporated into ARGONAUTE 4 (AGO4) and AGO6 proteins. Recently, non-canonical RdDM pathways have been discovered wherein sRNAs generated from Pol II transcripts (which were thought to silence TEs only post-transcriptionally) also direct DNA methylation.
Multiple non-canonical RdDM mechanisms have been reported in single locus studies; however, only recently have we investigated the role of all known RdDM pathways to silence TEs on the genome-wide level. By analyzing the features of TE targets for both canonical and non-canonical RdDM, I found key differences that distinguish the TE targets: the canonical 24nt sRNA mediated RdDM mainly targets the edges of TEs that are near genes, whereas the non-canonical 21-22nt sRNA mediated RdDM pathway (which utilizes Pol II mRNAs) causes methylation throughout the length of its target TEs. I also determined that full-length TEs, capable of self-transposing and/or catalyzing non-autonomous TE transposition, are preferentially targeted by 21-22nt sRNAs. This preference is driven by the selective cleavage of full-length TE mRNAs, which subsequently generates the secondary 21-22nt sRNAs that drive DNA methylation. The finding demonstrates that chromatin silencing patterns can be reflections of RNA-based degradation specificities. In addition, I discovered a category of Pol IV-independent 24nt sRNAs that also initiate DNA methylation. This pathway has a minor role in TE silencing in Arabidopsis, where TEs are silenced, but a much more significant role in the TE-rich maize genome. Together, these recent discoveries demonstrate that there is a complex network of sRNAs that synergistically target TEs for post-transcriptional degradation with the end result of chromatin modification.

P0565: Brassicas, Arabidopsis, and related
Tissue-Specific Divergence in the Alternative Splicing of Duplicated Gene Pairs in *Arabidopsis thaliana*
Grant William de Jong, University of British Columbia, Vancouver, BC, Canada

Whole-genome duplication events have played an extensive role in the evolution of flowering plants. The sudden doubling of genetic material can expedite large scale changes in gene function and expression patterns. Alternative splicing (AS) offers an avenue with which genes duplicated in polyploidy events (homeologs) may contribute to such functional diversity. AS produces multiple transcript isoforms through the differential removal of introns from the primary mRNA transcript. This process contributes to the function complexity of the cell by expanding both proteomic diversity and mechanisms of post-transcriptional regulation.
While there is evidence of considerable divergence in alternative splicing between homeologous genes, the extent to which these differences manifest in different tissue-types is less understood. Using RNA-Seq I have surveyed the transcriptomes of root tissues in paleopolyploid *Arabidopsis thaliana*. Assessing the distribution of splicing events between homeolog pairs, I am determining the extent of AS divergence between homeolog pairs across multiple zones of root development. I have identified tissue-specific alternative splicing profiles between several gene pairs. These results will hopefully provide insight into how alternative splicing contributes to tissue-specific variation following gene duplication in plants.

P0566: Brassicas, Arabidopsis, and related
Progress in Implementing Plastid Transformation in *Arabidopsis thaliana*
Pal Maliga[1], Qiguo Yu[1], Lisa Lamanna[1], Megan Kelly[1] and Kerry Lutz[2], (1)Rutgers University, Piscataway, NJ, (2)Farmingdale State College, Farmingdale, NY
We have reported high frequency plastid transformation in *Arabidopsis thaliana* based on hypersensitivity to spectinomycin in *ACC2* null mutants lacking a plastid-targeted acetylcoenzyme A carboxylase. As it was difficult to obtain fertile transplastomic plants in the hypersensitive Sav0 accession, we deleted the *ACC2* gene in the RLD and Ws accessions using the CRISPR/Cas9 system. RLD and Ws were chosen because they readily yield plants in tissue culture when exposed to plant growth regulators. Progress will be reported on plastid transformation in Arabidopsis with new vectors which do not impose a metabolic burden on the plants in *ACC2*-knockout RLD and Ws ecotypes.

P0567: Brassicas, Arabidopsis, and related
High-Throughput Screening Tools for Identification of Traits Contributing to Salinity Tolerance
Martin Trtlik[1], Klara Panzarova[2], Arianna Nigro[3], Jiri Fajkus[1], Zuzana Benedikty[1], Mark A. Tester[4] and Mariam Awlia[5], (1)PSI (Photon Systems Instruments), spol. s r.o., Drásov, Czech Republic, (2)Photon Systems Instruments (PSI), Drásov, Czech Republic, (3)Institute of Plant Biology, University of Zürich, Zürich, Switzerland, (4)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia, (5)Center for Desert Agriculture, KAUST, Thuwal, Saudi Arabia
Non-invasive capturing and interpreting of plant structural and functional phenotypes in controlled or dynamically changing environment is long-standing and necessary requirement for genetic and physiological research by crop breeders, agricultural industry, and academia. To sustain global food security the major challenge global agriculture and plant biology field has been facing is the identification of new high-yielding genotypes of agricultural crops that are adapted to our future climate. Soil salinity is one of the main stress factors that are severely affecting the agriculture land in global scale and results in significant reduction of plant growth and yield. It was shown that plants suffer a rapid growth reduction upon the first exposure of their roots to salt stress, which is occurring prior to the accumulation of ions to toxic concentrations in the shoots. During this early phase, symptoms of growth reduction include slower leaf emergence and a small growth size. The phenotypic traits associated with this type of tolerance can be quantified in the days immediately after imposition of stress using non-destructive image-based phenotyping.
To enhance our understanding of the early responses to salinity, we designed an experimental protocol based on using high-throughput and non-invasive imaging technologies developed at Photon Systems Instruments (PSI, Czech Republic). The methodology presented is based on automated integrative analysis of photosynthetic performance, growth analysis and color index analysis at the onset and early phase of salinity stress response of *Arabidopsis thaliana* ecotypes grown in soil. Here we show that stress imposition significantly and rapidly affected photosystem II operating efficiency, subsequently impacted growth dynamics and greening index of Arabidopsis plants at different stages of stress response.
Our work provides quantitative insights into early phase of salinity response and provides robust protocol for high-throughput image-based analysis of phenotypic traits associated with this early phase of salinity response. We show that the integrative concept of PlantScreen^{TM} high-throughput phenotyping platform provides a powerful tool for acquisition and selection of morphological, physiological and biochemical parameters, which can be used for identification of various components underlying early plant responses to various environmental conditions.

P0568: Brassicas, Arabidopsis, and related
QTL-Seq Analysis of Heat Tolerance in Broccoli
Sandra E. Branham, USDA-ARS, Charleston, SC and Mark W. Farnham, USDA-ARS Vegetable Laboratory, Charleston, SC
Broccoli (*Brassica oleracea* var. *italica*) production, worth approximately a billion dollars annually in the United States alone, is restricted in terms of location and season, due to the sensitivity of commercial cultivars to high temperatures. Heat stress during heading causes yield and quality loss of the crop. Several heat tolerant broccoli lines have been developed in recent years but few studies have examined the underlying genetic basis. Three years of summer field trials were used to evaluate heat tolerance in a doubled haploid population of broccoli. Bulked segregant analysis was combined with whole genome resequencing for a QTL-seq study of heat tolerance. Two QTL were identified and collocated with several strong candidate genes, including those encoding heat shock proteins and antioxidants.

P0569: Brassicas, Arabidopsis, and related
Whole Genome Resequencing of Canola for Variant Discovery and Routine Genotyping
Michelle Malmberg, Agriculture Victoria Research, Melbourne, Australia
Canola is the second most important oilseed crop worldwide. Although reduced representation GBS methods are valuable, whole genome resequencing (WGR) provides unbiased, genome-wide markers and can be used for the development of genomic resources. WGR of c. 150 canola varieties representative of global oilseed diversity was performed, and a list of c. 4 million validated and annotated SNPs orientated by

the reference genome was generated. This provides the basis of a variant database, increasing the reliability and repeatability of canola genomics including functional studies, as over 50% of genes were found to contain at least one non-synonymous substitution. The whole genome sequences were further used to evaluate and optimise the application of skim WGR, a method which is increasingly relevant as the cost of sequencing continues to fall. Skim coverage fastq files were generated *in silico* to represent a range of sequencing depths, and the effect of sequencing coverage and filtering depth on total SNPs, marker spread and genotyping accuracy was evaluated. As a broad recommendation, 1x coverage with stringent depth filtering was found to be cost-effective, providing c. 19K SNP markers, comparable to other GBS systems. Due to the flexible nature of the method, sequencing depth and filtering stringency can easily be tailored for individual sample sets and adjusted to produce the desired marker density. The SNP list developed here and the optimised skim WGR protocol will aid future genomics studies by allowing cost-effective and reliable genotyping across a range of populations.

P0570: Brassicas, Arabidopsis, and related
Sub-Pangenomic and Pan-Subgenomic Variation in the Recent Allopolyploid *Brassica napus*
Birgit Samans[1], Bjoern Usadel[2], Gunhild Leckband[3], Amine Abbadi[3] and Rod Snowdon[1], (1)Justus Liebig University, Giessen, Germany, (2)Forschungszentrum Jülich GmbH, Jülich, Germany, (3)NPZ Innovation GmbH, Holtsee, Germany
Genome fractionation after ancient polyploidisation events was a major driver of pangenome variation in plant species. Investigating gene presence-absence variation in the context of recent polyploidy can therefore provide valuable insight into crop evolution and genome dynamics. Genes duplicated via allopolyploidisation increase tolerance for structural rearrangements (e.g. homeologous exchanges and gene conversion events) or functional disruption through mutations (pseudogenisation, neofunctionalisation and subfunctionalisation). Both kinds of changes commonly lead to divergence in gene expression. We analysed the extent of pan-subgenomic variation in *Brassica napus*, a recent allopolyploid that rapidly became a major crop (rapeseed, canola). Genome and transcriptome sequences from 100 diverse, winter-type rapeseed accessions revealed transcript abundance profiles and structural genome variation in a sub-pangenomic context. A genome-wide survey of presence-absence variation in association with gene expression patterns enabled us to characterize the *B. napus* pangenome from a unique pan-subgenomic perspective, distinguishing different classes of core and dispensable genes across homeologous gene pairs representing the two *B. napus* subgenomes. Genes where the same single copy was always retained in the same subgenome were classified as "true core genes". The subset of these genes for which both subgenome homeologs were retained and expressed across all accessions were termed "core duplicates", a class of genes that might be essential for successful allopolyploidisation. Gene pairs for which either one or the other homeolog was retained in all accessions were classified as "subgenomic core genes". *Sensu stricto* pangenome terminology would in fact classify the latter as "dispensable" genes, however consideration of post-allopolyploidisation redundancy necessitates a refined definition of that term to consider homeologous gene pairs. Subgenomic core genes may indeed be individually dispensable, but perhaps only if their loss is compensated by retention of a closely related homeolog. On a species level, subgenomic core genes may contribute to enhanced adaptative capacity of polyploid plants and intraspecies hybrids.

P0571: Brassicas, Arabidopsis, and related
Genome-Wide Association Study of Important Traits in *Brassica napus* using Nested Association Mapping Population
Kede Liu and Jianlin Hu, Huazhong Agricultural University, Wuhan, China
Yield and yield-related traits are complex quantitative traits in crops. There are two general methods for dissecting the genetic basis of quantitative traits, which includes linkage analysis based on the segregating population derived from parents and genome-wide association study based on the natural population. Nested association mapping (NAM) population based on multi-parental hybridization has the advantages and overcomes defects of both segregating population and natural population, and has been widely applied to dissect the genetic basis of complex quantitative traits in maize. We constructed a NAM population in oilseed rape which derived from crosses between a common parent “Zhongshuang 11” and 15 diverse inbred lines. The NAM population contains 2,425 recombination inbred lines (RILs) from 15 segregating populations. Yield and yield-related traits including silique number, silique length, seeds per silique, seed size, branch number, length of main inflorescence, flowering time, plant height and oil content were investigated in multiple environments. Meanwhile, high-density genetic linkage maps were constructed for each RIL population using genotyping-by-sequencing. We integrated the 15 genetic linkage maps into a consensus linkage map. The genetic architecture of these yield and yield-related traits were dissected by both linkage and genome-wide association analysis. A large number of loci significantly associated with rapeseed yield and yield-related traits were identified. Causative polymorphisms underlying several loci were identified for yield and yield-related traits.

P0572: Brassicas, Arabidopsis, and related
GWAS Unravel Important Haplotype Regions in Hybrid Spring-Type Canola (*Brassica napus*) for Flowering Time and Seed Glucosinolate

P0573: Brassicas, Arabidopsis, and related
Homologous Pairing Control in *Brassica napus*
Erin Higgins[1], Wayne E. Clarke[1,2] and Isobel Parkin[1], (1)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (2)New York Genome Centre, New York, NY
Homologous recombination during meiosis is essential for creating genetic diversity and faithful separation of the chromosomes to produce haploid gametes. In the allotetraploid *Brassica napus* the presence of both the A and C genomes requires a process to restrict chromosome pairing and recombination so it occurs only between true homologues and not homoeologous chromosomes from the related genome. Identification of a genetic control mechanism and manipulation of pairing control genes in *B. napus* would be very valuable for crop improvement through the generation of novel copy number variation, or the introgression of genes from wild relatives, and the subsequent restoration of genetic stability that would be required to maintain these improvements.
A *B. napus* DH population segregating for control of homologous chromosome pairing and recombination was developed from a cross between an adapted *B. napus* line and a resynthesized line from a cross between *B. rapa* and *B. oleracea*. Individuals from this population were crossed

with an unrelated *B. napus* cultivar to produce testcross populations that were phenotyped for homologous recombination using the Brassica 60K Illumina SNP array. Using the SNP array to analyse testcross populations we were able to measure homoeologous recombination using reciprocal gain/loss of A and C genome SNP alleles and these data were used to map QTLs that control homologous recombination in *B. napus*. The alignment of SNP array probe sequences to the *B. napus* genome assembly allows for investigation of the regions underlying these QTLs and identification of candidate genes which control homologous chromosome pairing.

P0574: Brassicas, Arabidopsis, and related

Identification and Molecular Mapping of a Heterotic Locus in Canola (*Brassica napus* L.)

Puying Zheng[1], Yunming Long[1], Mukhlesur Rahman[1] and Wun S. Chao[2], (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

Heterosis exhibits superior performance for F1 hybrids generated by crossing two inbred parents and has been applied regularly in crop production. However, the molecular basis underlying heterosis remains largely unknown. In this study, we have identified a heterotic locus in the F_1 hybrids between 'Westar' (canola, *Brassica napus* L.) and its chromosome segment substitution line (CSSL) introgressed a chromosome segment in A10 of 'Surpass 400'. The F1 hybrids, 'Westar' × CSSL and CSSL × 'Westar', increased 7% and 20% in grain yield compared to the best-parent value respectively under the growth chamber conditions. To further map the heterotic gene, a set of new homozygous introgression lines (ILs) were developed carrying different lengths of introgressed chromosome segments in A10 of 'Surpass 400'. Eight ILs were crossed with Westar reciprocally. The resultant 16 F1 hybrids along with their parents (ILs and Westar) were planted in two locations (Fargo and Prosper, ND) using a randomized complete block design with three replicates to estimate heterotic effects. Based on grain yield of 16 hybrids (compared to the best parent value) and genotype data (position and length of introgressed Surpass 400 chromosome segments), we have mapped the heterotic locus within the 500 kb region.

P0575: Brassicas, Arabidopsis, and related

Analysis of Transcriptional and Epigenetic Changes in Hybrid Vigor of Allopolyploid *Brassica napus* uncovers Key Roles for Small RNAs

Yifei Shen, Institute of Crop Science, Zhejiang University, Hanghou, China, Daguang CAI, Christian-Albrechts University of Kiel, KIEL, Germany, Micheal Timko, University of Virginia, Charlottesville, Virgin Islands (U.S.), Qian-Hao Zhu, CSIRO Agriculture Flagship, ACT, Australia and **Longjiang Fan**, Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou, China

Heterosis is a fundamental biological phenomenon characterized by the superior performance of a hybrid compared to its parents. The underlying molecular basis for heterosis, particularly for allopolyploids, remains elusive. In this study, we analyzed the transcriptomes of Brassica napus parental lines and their F1 hybrids at three stages of early flower development. Phenotypically, the F1 hybrids show remarkable heterosis in silique number and grain yield. Transcriptome analysis revealed that various phytohormones (auxin and salicylic acid) response genes are significantly altered in the F1 hybrids relative to the parental lines. We also found evidence for decreased expression divergence of the homoeologous gene pairs in the allopolyploid F1 hybrids and suggest that high-parental expression level dominance plays an important role in heterosis. Small RNA and methylation studies aimed at examining the epigenetic effect of the gene expression level changes in the F1 hybrids showed that the majority of the small interfering RNA (siRNA) clusters had a higher expression level in the F1 hybrids than parents, and that there was an increase of the genome-wide DNA methylation in the F1 hybrid. Transposable elements associated with siRNA clusters had a higher level of methylation and a lower expression level in the F1 hybrid, implying that the non-additively expressed siRNA clusters resulted in lower activity of the transposable elements through DNA methylation in the hybrid. Our data provide insights into the role that gene expression pattern changes and epigenetic mechanisms contribute to heterosis during early flower development in allopolyploid B. napus.

P0576: Brassicas, Arabidopsis, and related

eQTL-Guided Co-Expression Analysis for Constructing Regulatory Network using a Synthetic *Brassica napus* F2 Population

Ruijuan Li[1], John Davis[1], Kwang-Ju Jeong[2], Sunboung Lee[3], Seungmo Kim[1], Shinje Kim[2], Richard Michelmore[4] and Julin N Maloof[1], (1)University of California, Davis, CA, (2)Fungi and Plants Co.,Ltd, Chungcheongbuk-do, Korea, Republic of (South), (3)Fungi and Plants Co., Ltd, Chungcheongbuk-do, South Korea, (4)Genome Center, University of California, Davis, Davis, CA

Brassica napus (AACC), an economically important crop, is an allotetraploid species that resulted from hybridization between two diploid species *B. rapa* (AA) and *B. olereacea* (CC) followed by chromosome doubling. We have synthesized two new *B. napus* genotypes Da-Ae (AACC) and Da-Ol-1 (AACC). These strains differ in fatty acid content and the level of drought tolerance. While Da-Ae was selected from the synthetic allopolyploid F2 population created by hybridizing *B.rapa* with *B.olereacea*, Da-Ol-1 was made by crossing *B.napus* with *B.juncea* (AABB) and then backcrossing to *B.napus*. Making use of the genotyping information generated from RNA-seq data of the two lines and their F2 mapping population of 166 lines, we constructed a genetic map consisting of 2,013 SNP markers that spans 2,884 cM across 19 linkage groups. eQTL analysis with this genetic map identified 728,717 eQTL for 21,823 genes, comprising 492,708 (67.6%) cis-eQTL and 236,009 (32.4%) trans-eQTL. Co-expression and network analysis using these eQTL and expression data would help us identify modules revealing association between transcript level patterns and biological processes, such as fatty acid content levels and flowering time, etc. By integrating eQTL and QTL analysis results, we would be able to identify potential causative genes for various biological traits.

P0577: Brassicas, Arabidopsis, and related

Quality Breeding and Its Hidden Effect on Seed Germination Performance in Winter Oilseed Rape (*B. napus* L.)

Sarah Hatzig, Department of Plant Breeding, Justus Liebig University, Giessen, Germany and Rod Snowdon, Justus Liebig University, Giessen, Germany

Rapeseed is one of the most important oil crops worldwide. Intensive breeding efforts have led to a broad usability, not least because specific seed qualities – low glucosinolate content and zero erucic acid quality - could be established in the last half century. While main effort is done

on the improvement of yield quantity and quality, secondary agronomic traits, amongst them seed vigor were often less valued in the modern quality germplasm collection. The great agronomic and economic significance of optimized seed vigour and germination was underlined by the fact that it is closely linked to population density, uniformity, compensation propensity and the required sowing quantity. Along with all efforts to directly enhance crop performance, it should therefore not be neglected that unintentional co-selection of unconsidered secondary traits could also curtail the breeding progress. The present study underpins this assertion, as it demonstrates that breeding for specific seed qualities in winter oilseed rape has already implicated a restriction in seed germination performance. In a test set comprising 215 diverse winter oilseed rape varieties, low seed erucic acid content and reduced seed glucosinolate content were significantly related with inferior seed germination. This finding emphasizes that breeding should be conducted in a more holistic way, integrating agronomic traits - such as seed germination - with a fundamental importance for adequate growth performance and yield building. Furthermore the current study delivers a practical approach for the selection of improved seed germination within modern quality pools. Within a genome-wide association approach, we identified partially overlapping QTL modulating germination speed as well as seed quality. In this context we elaborated, that germination performance in the modern quality gene pool could be improved by marker assisted selection, without disturbing seed quality performance.

P0578: Brassicas, Arabidopsis, and related

Secondary Dormancy Potentials of a Diverse Collection of *Brassica napus* L. Lines Grown in Different Environmental Conditions and the Relationship with Seed Vigour Parameters

Caroline H. Brown, Agriculture and Agri-foods Canada/University of Saskatchewan, Saskatoon, SK, Canada, Sally Vail, Agriculture & Agri-Food Canada, Saskatoon, SK, Canada and Steve Shirtliffe, University of Saskatchewan, Saskatoon, SK, Canada

Canola (*Brassica napus* L.) seed which germinates after the cropping season the seed was intentionally planted is referred to as volunteer canola and ranks 4th in most occurring weed on the Canadian prairies. Volunteer canola is becoming increasingly hard to control as more herbicide tolerant crops are grown. Secondary dormancy is the physiological mechanism leading to the extended presence of seed in the seedbank. Secondary dormancy is inducible dormancy after the seed is released from the mother plant and due to conditions including low light, temperature and moisture.

A diverse collection of 51 spring *B. napus* lines (Nested Association Mapping founder lines) was phenotyped for secondary dormancy and seed vigour when grown under four environmental conditions (two Canadian and two Chilean). Direct selection for lines with low secondary dormancy has not been performed before therefore; the effect on seed vigour is unknown. It is hypothesized that a wide range in dormancy potentials exist across lines and environments as well as a negative relationship between secondary dormancy and seed vigour parameters. Secondary dormancy potentials across lines are significantly different ranging from 0 to 84% dormant (p-value <0.0001) as well as a significant line by environment interaction (p-value <0.0001). Low dormancy lines are more consistent across environments than higher dormancy lines. The seed vigour test, controlled deterioration (CDT) showed a non-significant correlation (p-value= 0.45) between the CDT and secondary dormancy , suggesting the two are not associated and selecting for low dormancy genotypes without compromising seed vigour is achievable.

P0579: Brassicas, Arabidopsis, and related

Transcriptome Analysis Suggests Cytokinin and Brassinosteroid Signalling may account for Differences between Spring and Winter Canola (*B. napus*) Root Development

Muhammad Arifuzzaman[1], David P. Horvath[2] and Mukhlesur Rahman[1], (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

Winter type canola produces significantly larger and more vigorous roots compared to spring type canola at late growing stage. To understand the mechanism, we analyzed transcriptomes from the root samples collected from two spring and two winter types at 60 days. A total of 169,646 *Brassica* gene models were analyzed and 1474 gene models were found to be significantly differentially expressed between spring and winter types. Auxin is known to significantly impact root growth and development. Although a few auxin responsive family genes were significantly differentially expressed between spring and winter types, there was no distinct pattern of differential regulation observed in either of the growth types. This indicates some other non-auxin mediated root growth regulation might be responsible for the root system between the two types. *ARABIDOPSIS RESPONSE REGULATOR (ARR)* genes involved in cytokinin responsive signaling pathway and *ARGOS-like* genes involved in brassinosteroids mediated signaling pathway were significantly differentially expressed between spring and winter types. Higher expression of ARR genes in spring type suggests the presence of higher concentration of cytokinin which has a proven inhibitory effect on root growth. Gene set enrichment analysis also suggests that gene sets related to cytokinin signaling, cytokinin metabolism and ARR-A type family were upregulated in spring type. In addition, gibberellin responsive, gibberellin signaling gene sets were also upregulated in spring types. Gibberellin may prevent lateral root development by inhibiting auxin carrier *Pin-formed 9 (PIN9)*. Extensive gene expression data generated in this research will further assist to understand the natural variation of root system in canola growth habits.

P0580: Brassicas, Arabidopsis, and related

Breeding Better Crops: Investigating the Genetic Basis of Micronutrient Efficiency in *Brassica napus*

Aoife Grace Sweeney, University of York, York, United Kingdom and Ian Bancroft, Centre for Novel Agricultural Products (CNAP), York, United Kingdom

Unlike deficiencies in macronutrients (such as N/P/K), micronutrient deficiencies are capable of significantly reducing crop yield prior to the observation of typical deficiency symptoms. Given the growing political, environmental and monetary pressures on crop production, investigating how genetic variation determines susceptibility to deficiency will be essential in ensuring stable crop yields. Our research focuses on exploiting underlying natural genetic variation in *Brassica napus* (Oilseed rape/Canola) to explore micronutrient efficiency mechanisms. Using an Associative Transcriptomic approach, ionomic differences between plants from a large diversity panel were compared for gene sequence and gene expression variation. From these results, it was possible to identify candidate genes which could be further tested through orthologous genes in *Arabidopsis thaliana* with T-DNA insertional mutants. Once the role of candidate genes in micronutrient efficiency were

verified in *A. thaliana,* markers from Associative Transcriptomic outputs could be used in marker assisted selection to improve micronutrient use efficiency. These improvements will enable the stabilisation of crop yield and a reduction in fertiliser inputs.

P0581: Brassicas, Arabidopsis, and related

Investigating the Genetic Basis of Boron Efficiency in *Brassica napus* (*B. napus*)

Helen Elizabeth Riordan, University of York, CNAP, York, United Kingdom and Ian Bancroft, Centre for Novel Agricultural Products (CNAP), York, United Kingdom

Globally, of the arable land under permanent cropping, millions of hectares are susceptible to nutrient deficiencies. Boron-deficiency is a particular issue, with the model utilised within this project, the crop plant *B. napus,* being particularly susceptible. Numerous deficiency disorders have also been characterised, including 'brown heart' in the *B. napus* cultivar swede. Further understanding into the genetic basis of boron-efficiency would therefore be beneficial, allowing for both a reduction in fertiliser-input and the maximisation of crop yields.
Using an 'Associative Transcriptomic' (AT) approach, utilising a diverse panel of *B. napus* lines, this project aims to elucidate how variation in both the nutrient and genetic components of plants are linked. The use of AT enables us to assess how variation in both gene sequence and expression, in the form of single nucleotide polymorphisms (SNPs) and gene expression markers (GEMs) respectively, affects boron phenotype. Using this approach, candidate genes can be further analysed in *Arabidopsis thaliana*, itself a member of *Brassicaceae* family, which exhibits extensive genome collinearity with *B. napus.* Utilising *Arabidopsis* T-DNA insertions lines to knock-out chosen genes, the ionome (total mineral nutrient and trace element composition) of these mutants can be studied. Any observed disruption in the boron content of these mutants is indicative of a role within boron-homeostasis, and hence validates the candidate gene. Through the use of AT, it should be possible to identify underlying genetic factors controlling boron-use, thus enabling marker-assisted selection for increasingly nutrient-efficient crops.

P0582: Brassicas, Arabidopsis, and related

Dissection of the Genetic Architecture of Three Seed-Quality Traits and Consequences for Breeding in *Brassica napus*

Bo Wang[1], Zhikun Wu[1], Qinghua Zhang[1], Yingjie Xiao[1], Jiangsheng Wu[1], Graham J. King[2], **Haitao Li**[1] and Kede Liu[1], (1)Huazhong Agricultural University, Wuhan, China, (2)Southern Cross Plant Science, Southern Cross University, Lismore, Australia

Genome-wide association studies (GWASs) combining high-throughput genome resequencing and phenotyping can accelerate the dissection of genetic architecture and identification of genes for complex traits in plant. In this study, we developed a rapeseed genomic variation map consisted of 4,542,011 SNPs and 628,666 INDELs. GWAS were performed for three seed-quality traits, including erucic acid content (EAC), glucosinolate content (GSC) and seed oil content (SOC) using 3.82 million polymorphisms in an association panel. A total of 6, 49 and 17 loci were detected to be associated with EAC, GSC and SOC in multiple environments, respectively. The average total contribution of these loci in each environment for EAC and GSC was 94.1% and 87.9%, which was much higher than that for SOC (40.1% in average). Resolution of these three traits across the genepool, which mainly inherited in an additive manner, will make the breeding more straightforward via pyramiding these loci. Furthermore, we found candidate genes underlying associated loci, followed by correlating the sequence variation with phenotypic variation based on the functional polymorphisms in gene regions. Two common *fatty acid elongase 1* (*FAE1*) genes at two major loci for EAC on chromosomes A8 and C3 and three common *MYB28* genes at three major loci for GSC on chromosomes A9, C2 and C9 were all found. Moreover, we detected correlation between GSC and SOC and sequence variation in four additional new genes. The present study provides insights into the genetic architecture of three seed-quality traits, which would be useful for genetic improvement of *B. napus*.

P0583: Brassicas, Arabidopsis, and related

Exploring the Genes Controlling Glucosinolate Variation in Oilseed Rape (*Brassica napus*)

Varanya Kittipol, University of York, York, United Kingdom

There is a great interest in breeding new varieties of agriculturally important *Brassica* oil crop with desirable glucosinolate contents, for instance high concentrations in the leaves for resistance against pests and low seed concentrations to reduce anti-nutritional substance in the meal. In order to achieve such aim, further understanding of the genetic basis is required to study the underlying mechanism that control quantitative variation in the glucosinolate profiles. To overcome the lack of reference genome sequences in *Brassica napus* and the difficulties of studying the complex polyploidy genomic, a modified GWAS method termed 'associative transcriptomics' (AT) has been employed. AT uses transcriptome sequencing to allow the discovery of single-nucelotide polymorphism (SNP) markers in tight linkage disequilibrium with causative genes as well as finding genes with expression patterns (gene expression markers, GEM) that correlated with the trait variation. In this study, AT has been employed to a panel of 300 *Brassica napus* accessions to identify genetic regions which are likely to be involved in controlling glucosinolate natural variation in the leaves and roots. Recent completion of the glucosinolate profiles for the whole diversity panel reveals intriguing relationships between the glucosinolate structural groups. Different glucosinolate patterns are observed, aliphatic glucosinolates are predominated in leaves whereas aromatic glucosinolates are common in roots, suggesting different regulation mechanisms between above- and below- ground organs. Initial AT analysis reveals several loci associated with high glucosinolates content, these loci are potentially controlling the variation of seed and leaf aliphatic glucosinolates in *Brassica napus*.

P0584: Brassicas, Arabidopsis, and related

Comparative Genome Analysis of Blackleg Resistant and Susceptible Genes in *Brassica napus* using Whole Genome Re-Sequencing

Ting Xiang Neik, Martin J. Barbetti, Philipp E. Bayer, Chon-Kit Kenneth Chan, Anita Severn-Ellis, David Edwards and Jacqueline Batley, University of Western Australia, Perth, Australia

Brassica napus (oilseed rape/canola), is highly valued for its oil and is widely grown in Europe, China, India, North America and Australia. However blackleg disease, caused by the fungal pathogen *Leptosphaeria maculans,* remains an ongoing, serious threat to canola crops in these regions. Breeding resistant *B. napus* varieties against this pathogen is more sustainable and cost-effective than conventional chemical control

approaches. Single major race-specific *R* genes are an excellent genetic resource enabling breeding of resistant varieties. These *R* genes display specific interaction with the *Avr* genes in *L. maculans* where recognition between the dominant *R* gene of the host and the corresponding dominant *Avr* gene of the pathogen results in a hypersensitive response and minimised host disease reaction. We assessed the host phenotype responses of 17 different *B. napus* lines, each carrying putative *Rlm* genes, inoculated with a set of differential *L. maculans* isolates, each isolate carrying one or more defined *Avr* genes. To determine the candidate genes responsible for either a hypersensitive/highly resistant or a susceptible reaction, we utilized a whole genome re-sequencing approach. SNP and gene presence/absence analysis was utilised to associate phenotype with genotype. The identified candidate genes can be utilised for breeding resistant *B. napus* cultivars.

P0585: Brassicas, Arabidopsis, and related

***Brassica napus* × *Brassica nigra* Hexaploid Hybrids to Introgress Blackleg Resistance into Rapeseed**

Roman Gaebelein, Justus Liebig University Giessen, Giessen, Germany and Annaliese S Mason, Justus Liebig University, Giessen, Germany

Leptosphaeria maculans is one of the main pathogens of rapeseed (*Brassica napus*) causing annual yield losses between 5 and 20% in the main growing areas of Europe, Canada and Australia. Applied resistance mechanisms are under constant pressure and frequently overcome by continuously evolving pathotypes.

Brassica nigra, a close relative of *B. napus*, is a potential donor for introgression of new resistance genes into oilseed rape.

Hybrids were produced by crossing male sterile and male fertile cultivars of *Brassica napus* with three different cultivars of *B. nigra*. Embryo rescue was applied to ensure the development of triploid hybrids. Hybrid plants were clearly distinguishable by intermediate phenotypic characteristics and the triploid karyotype was confirmed by flow cytometry.

Clones of 42 different F1 hybrid plants were subjected to in-vitro colchicine treatment, resulting in the development of 16 different hexaploid hybrid plants. Novel *Phoma* resistance in the *B. nigra* parents was confirmed by cotyledon inoculation tests using 11 different field-collected isolates from Germany.

Adult plant resistance status of the hybrids will be evaluated and suitability of triploid and hexaploid material for disease resistance transfer will be assessed by cytogenetic analysis of the first meiosis. GISH staining using *B. nigra* DNA will reveal the frequency of homeologous recombination with the B genome for physical evidence of gene transfer between the subgenomes. In future, this material is expected to provide a novel source of blackleg disease resistance for rapeseed breeding.

P0586: Brassicas, Arabidopsis, and related

LepR3/Rlm2* Genes: Resistance to *Leptosphaeria maculans* in *Brassicas

Hua Yang[1,2], Angela Van de Wouw[3], David Edwards[1,2] and Jacqueline Batley[2], (1)University of Queensland, Brisbane, Australia, (2)University of Western Australia, Perth, Australia, (3)University of Melbourne, Melbourne, Australia

Brassica species are at high risk of severe crop loss due to pathogens, especially *Leptosphaeria maculans* (the causal agent of blackleg). To date, many *Brassica R*-genes to *L. maculans* have been genetically defined, but only *Rlm2* and *LepR3*, allelic variants of the *LepR3/Rlm2* blackleg resistance locus on chromosome A10, have been cloned from *Brassica napus*. *LepR3* confers resistance to *L. maculans* avirulence gene *AvrLm1*, while *Rlm2* interacts with *AvrLm2*. In addition, *Rlm1*, located on chromosome A07, also conveys race-specific resistance to *AvrLm1*. The identification and characterization of host resistance genes in *Brassica* species, along with the corresponding pathogen avirulence genes will provide an insight into the evolution of disease resistance mechanisms and permit the development of novel approaches for sustainable canola production.

In this study, *LepR3/Rlm2* was amplified and sequenced in many *B. napus* cultivars. Fifty-four alleles were isolated from these cultivars. The amino acid sequences of these allelic genes contain a conserved N-terminal domain, C-terminal and LRR repeats. The main difference between them is the number of LRR repeats. We demonstrated that *Rlm1* and *LepR3* were different resistance genes conferring resistance to the same avirulence gene *AvrLm1*.

P0587: Brassicas, Arabidopsis, and related

Development of EST-SSR Markers in Flowering Chinese Cabbage (*Brassica campestris* L. ssp. *chinensis* var. *utilis* Tsen et Lee) Based on *De Novo* Transcriptomic Assemblies

Jingfan Chen[1], Ronghua Li[1], Yanshi Xia[1], Guihua Bai[2], Kadambot Siddique[3] and **Peiguo Guo**[1], (1)Guangzhou University, Guangzhou, China, (2)Hard Winter Wheat Genetics Research Unit, USDA-ARS, Manhattan, KS, (3)University of Western Australia, Perth, Australia

Flowering Chinese cabbage is one of the most important vegetable crops in southern China. Genetic improvement of various agronomic traits in this crop is underway to meet high market demand in the region, but the progress is hampered by limited number of molecular markers available in this crop. This study aimed to develop EST-SSR markers from transcriptome sequences generated by next-generation sequencing. RNA-seq of eight cabbage samples identified 48,975 unigenes. Of these unigenes, 23,267 were annotated in 56 gene ontology (GO) categories, 6,033 were mapped to 131 KEGG pathways, and 7,825 were assigned to clusters of orthologous groups (COGs). From the unigenes, 8,165 EST-SSR loci were identified and 98.57% of them were 1–3 nucleotide repeats with 14.32%, 41.08% and 43.17% of mono-, di- and tri-nucleotide repeats, respectively. Fifty-eight types of motifs were identified with A/T, AG/CT, AT/AT, AC/GT, AAG/CTT and AGG/CCT the most abundant. The lengths of repeated nucleotide sequences in all SSR loci ranged from 12 to 60 bp, with most (88.51%) under 20 bp. Among 170 primer pairs were randomly selected from a total of 4,912 SSR primers we designed, 48 yielded unambiguously polymorphic bands with high reproducibility. Cluster analysis using 48 SSRs classified 34 flowering Chinese cabbage cultivars into three groups. A large number of EST-SSR markers identified in this study will facilitate marker-assisted selection in the breeding programs of flowering Chinese cabbage.

P0588: Brassicas, Arabidopsis, and related

Resolving Genomic Interactions between Seed Storage Proteins and Glucosinolate in Mustard

Priyakshee Borpatragohain, Terry J Rose, Bronwyn Barkla and Graham J. King, Southern Cross Plant Science, Southern Cross University, Lismore, Australia

Oleriferous brassicas store reduced sulphur (S) in mature seeds mainly as secondary metabolite-glucosinolate (GSL) and storage proteins-cruciferin and napin. Glucosinolates have a wide range of positive aspects in food production, human nutrition and plant defence. The globulin-like cruciferins are a potential source of high quality vegetable proteins for human and livestock consumption. However, anti-nutritional effects of GSLs on livestock, and allergenicity of napins in humans, hinders the use of protein-rich brassica seed meal. We have recently proposed a model for interaction between S availability and GSL content in the context of source-sink relationships during crop development, which can be extended to account for S flux in the seed sink affecting protein and GSL ratios. There have been some reports of an interaction between storage proteins and GSL content in canola (*B. napus*) seeds. However, few studies have been carried out to resolve the genomic interactions between GSL and storage proteins in mustard (*B. juncea*). Indian mustard has advantages over canola and other cultivated species of brassicas due to inherent drought tolerance, pod shattering and nematode resistance, as well as the ability of rapid and vigorous ground cover to reduce weed growth. We are carrying out a QTL study with a double haploid mapping population of Indian mustard having parents of low- and high-GSL, to understand how cruciferin and napin content varies in relation to seed GSL content. A relatively higher abundance of proteins is detected using 1D-SDS PAGE in the low-GSL parent compared with the high-GSL parent, substantiated by Western blot analysis using antibodies against seed storage proteins. We are currently resolving QTLs for S, seed storage proteins and GSL content to refine our proposed model focusing on manipulating GSLs as S sink by altering the amount of seed storage proteins.

P0589: Brassicas, Arabidopsis, and related

Effect of the Genotype and Nutrient Medium on Microspore Cultures Microspore Derived Embryos in *Brassica juncea*

Ravinder Grewal[1], Iresha Alahakoon[2], Bifang Cheng[1], Murray Lewis[1], Eric Johnson[3] and Christina Eynck[1], (1)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (2)Mustard-21 Inc., Saskatoon, SK, Canada, (3)Deaprtment of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

Brassica juncea is an important oilseed crop in many countries including Canada, gaining popularity for arid zone cultivation. Doubled haploid (DH) technique has played an important role in shortening the breeding cycle but success of this technology has been limited in oilseed crops like *B. juncea* due to low embryo yield. This study involves establishing an efficient and reliable production of embryogenesis in *B. juncea* by studying the influence of different factors; genotype and nutrient medium on microspore embryo formation. Microspores from six genotypes of *B. juncea* were cultured on different modifications of the NLN medium. Genotype and medium dependence for embryo formation was observed. Significant differences in the response of genotypes, culture medium and genotype x medium interactions were observed for microspore yield, higher response for androgenic embryo formation and haploid plant regeneration. The majority of plants were obtained from microspores isolated from buds of length 2.5 to 3.5 mm and cultured in NLN liquid medium with 17% sucrose (w/v) supplemented with 24-epibrassinolide and 0.5mg/L BAP (6-benzylaminopurine) during first 48 hours and changed to 10% (w/v). Embryos cultured on B5 medium with 2% sucrose (w/v), 0.1GA3 and 1% agar stimulated the large number of haploid plant formation. This protocol is in use for the selection of herbicide tolerant lines.

P0590: Brassicas, Arabidopsis, and related

Genome-Wide Analysis of NBS-LRR Genes in Indian Mustard (*Brassica juncea*) and Prediction of Candidate Disease Resistance Genes

Fabian C. Inturrisi[1], Philipp E. Bayer[1], Hua Yang[1,2], Chon-Kit Kenneth Chan[1], David Edwards[1] and **Jacqueline Batley**[1], (1)University of Western Australia, Perth, Australia, (2)University of Queensland, Brisbane, Australia

Brassica juncea (Indian mustard), an annual crop, is a member of the economically important Brassicaceae family and commonly cultivated as either brown or oriental mustard. There are a number of important diseases affecting *B. juncea*, including blackleg (*Leptosphaeria maculans*) and white rust (*Albugo candida*). For crop improvement, it is important to identify resistance (R) genes for disease resistance to be used in breeding. Availability of the *B. juncea* genome has aided in genome-wide analysis of nucleotide binding site leucine-rich repeats (NBS-LRR) genes, an important class of R genes. A total of 289 NBS-LRR genes were identified in *B. juncea* and the phylogenetic relationship of these NBS-LRR genes was analysed. NBS-LRR genes were predominately TIR type and 45% of genes were clustered. Published QTL for disease resistance in *B. juncea* were investigated where their physical position were determined, and NBS-LRR genes located within the region were investigated as candidate R genes. Nineteen candidate NBS-LRR genes were identified for further analysis for blackleg and white rust resistance. Genome-wide analysis of *B. juncea* NBS-LRR provides an important resource to identify candidates for novel and functional R genes, which is a key step to enhance resistance against pathogens.

P0591: Brassicas, Arabidopsis, and related

Creating New *Brassica oleracea* × *B. juncea* Allohexaploids through Ovule Culture

Margaret W. Mwathi, University of Queensland, Brisbane, Australia, Jacqueline Batley, University of Western Australia, Perth, Australia and Annaliese S Mason, Justus Liebig University, Giessen, Germany

The *Brassica* genus contains a large number of diploid and allotetraploid crops, including oilseeds, vegetables and condiments. The close relationship between *Brassica* species enables interspecific hybridisation to be used for crop improvement: one new breeding goal is production of a diverse allohexaploid crop species.

Hand pollination between 2 genotypes of *B. juncea (*Xingyou 4 and B578*)* (2n=AABB) with *B. oleracea* (TO1000) and wild C genome species *B. incana*, *B. montana* and *B. cretica* were performed (752 total bud pollinations, average 57.8 per cross combination) in both cross directions. The combination with *B. oleracea* was highly successful with a crossabilty ratio of 0.125 (number of plants in final multiplication media/ the total number of flowers pollinated), producing 7 triploid hybrids (2n=ABC) from 85 flowers and 35 cultured ovules. Ovule rescue was used to overcome hybridization barriers.

Pollen fertility was low; between 2–10% (average fertility 5.8%). Ploidy confirmation from flow cytometry, phenotype and meiotic analysis confirmed hybrids as 3x triploids, and no pods or seed setting was observed in any hybrid. Confirmed (3x) hybrids were multiplied and treated

with colchicine of varying concentrations (0.5%, 0.1%, 0.15%, 0.2% and 0.25% w/v; 8 cuttings/group). Changes were seen in pods, leaves and stems in all groups; however, no such changes were in untreated control. All colchicine treatment categories produced seeds, with the 0.15% and 0.2% groups producing the most (62 and 58 seeds respectively). A total of 200 seeds were harvested and will be sown for further analysis of meiotic stability and chromosome inheritance.

P0592: Brassicas, Arabidopsis, and related

Demographic History of Morphotype Diversification in *Brassica oleracea*

Sarah D. Turner[1], Makenzie E. Mabry[1], J. Chris Pires[1] and Timothy M. Beissinger[2], (1)Division of Biological Sciences, University of Missouri, Columbia, MO, (2)USDA-ARS, University of Missouri, Columbia, MO

The vegetables encompassed by *Brassica oleracea*, commonly known as cole crops, can be categorized into six distinct and diverse morphotypes: kale, cabbage, Brussels sprouts, kohlrabi, broccoli, and cauliflower. These crops are an especially interesting model for domestication; wild mustard, a small, unpalatable plant, was selected for a wide range of appearances, flavors, and nutritional attributes. Despite widespread use of these crops to demonstrate the power of human selection on crop domestication, the demographic history of *B. oleracea* remains poorly understood. Using resequenced genomes from the major morphotypes in *B. oleracea*, we present a demographic analysis to estimate ancestral population sizes and population splits. Results from this work will expand our understanding of how humans have impacted the evolution of cole crops, providing insight into Brassica evolution and facilitating future crop improvement efforts.

P0593: Brassicas, Arabidopsis, and related

Uncovering the Effects of Sequence Polymorphisms, Gene Expression, and Alternative Splicing in Different Crops of *Brassica oleracea*

Po-Xing Zheng, Jheng-Yang Ou and **Yao-Cheng Lin**, Biotechnology Center in Southern Taiwan, Agricultural Biotechnology Research Center, Academia Sinica, Tainan, Taiwan

Nucleotide sequence polymorphisms and alternative splicing (AS) increase the proteome diversity by using different combination of exons and intron retention from the same genomic loci. In *Brassica oleracea*, many popular crops, including broccoli, cabbage, cauliflower, kale, Kohlrabi, Brussels sprouts, and collards, have versatile of morphology diversity, but the underlining mechanisms of the phenotypic plasticity is still unknown. In this study, we hypothesized that the sequence polymorphisms, changes of gene expression values and AS on flowering-related genes have contributed to the domestication of different crops in *B. oleracea*. To identify the sequence polymorphism, resequencing of 20 different cauliflower genomes were obtained from BioProject PRJNA312457. Through analysis by GATK and SnpEff, single nucleotide polymorphisms (SNPs) were significantly accumulated in multi-copied genes, while insertions and deletions (INDELs) were found in single-copy genes. Furthermore, single-copy genes accumulated more mutations in splicing related regions, but flowering-related genes had higher frequency to alter protein sequence. To further analyze the effect of gene expression and AS on crop breeding of *B. oleracea*, a total of 91 RNA sequence data were obtained from BioProject PRJNA289196, including root and leaf transcriptomes of 7 types of *B. oleracea* crops. Gene expression were estimated by Salmon and co-expression network. Results showed that tissue type majorly contributed to distinguish different samples, while different crops shared similar transcriptomic patterns. Further analysis will focus on gene- and isoform-specific transcriptome difference. We will further extend the analysis to tropical *B. oleracea* to uncover novel mechanisms during domestication and contribute to molecular breeding in *B. oleracea*.

P0594: Brassicas, Arabidopsis, and related

Novel Unilateral Incompatibility in *Brassica rapa* Is Regulated by Duplicated Self-Incompatibility Genes, *PUI1* and *SUI1*

Yoshinobu Takada[1], Kohji Murase[2], Hiroko Shimosato-Asano[3], Takahiro Sato[1], Honoka Nakanishi[4], Keita Suwabe[5], Kentaro K. Shimizu[6], Yong Pyo Lim[7], Seiji Takayama[2], Go Suzuki[4] and **Masao Watanabe**[1], (1)Tohoku University, Sendai, Japan, (2)The University of Tokyo, Tokyo, Japan, (3)Nara Institute of Science and Technology, Ikoma, Japan, (4)Osaka Kyoiku University, Kashiwara, Japan, (5)Mie University, Tsu, Japan, (6)University of Zurich, Zurich, Switzerland, (7)Chungnam National University, Daejeon, South Korea

Most of flowers pollinated by insects prevent the undesirable pollen from different species. Within species, for maintain the genetic diversity, several outcrossing systems, dioecy, dichogamy and self-incompatibility (SI), etc., are established. About half of plants have SI trait. The SI is defined as the inability of a fertile hermaphrodite plant to produce zygotes after self-pollination. In *Brassica* species, this SI trait is sporophytically regulated by a single locus with *S*-multiple alleles. To date, SP11 (small cysteine-rich protein) and SRK (receptor-type protein kinase) have been identified as the male and female *S* determinants, respectively. These molecules have a role in the pollen-pistil interaction of SI. In the case of crossing between the plants with the same *S*-alleles, incompatibility phenotype is observed.

We have already identified novel unilateral incompatibility (UI) phenomenon between Japanese and Turkish lines within *B. rapa*. Even if combination between plants with different *S*-alleles, pollination of Turkish pollen on Japanese pistil leads to UI, whose phenotype is similar to SI.

In order to dissect this UI, we performed genetic mapping and genome analysis. A set of genes homologous to *SP11* and *SRK* was identified at the *UI* locus, which might have been duplicated from the different chromosome. PUI1 and SUI1, which were demonstrated as pollen and pistil determinants, respectively, should trigger the UI between Japanese and Turkish lines. In this presentation, we will show the functional and phylogenetic analyses of this UI.

P0595: Brassicas, Arabidopsis, and related

Effect of Knocking out *Bna.FAD2 Family* on the Composition of the Industrial Rapeseed Oil

Harjeevan Kaur and Ian Bancroft, Centre for Novel Agricultural Products (CNAP), York, United Kingdom

Brassicas used for the seed oil production are referred as “rapeseed” and its fatty acid composition defines the use. Oil from High Erucic acid rapeseed in Low PUFA (poly unsaturated fatty acid) background (abbreviated as HELP) has wide range of industrial uses. This oil has additional advantage of higher stability due to their low PUFA content. *Bna.FAD2 (Fatty acid desaturase)* and *Bna.FAE1 (Fatty acid elongase)*

are the main loci controlling the amount of PUFAs and erucic acid in the seeds of *Brassica napus*, respectively. High erucic acid cultivars were cross pollinated to the lines with knocked out *Bna.FAD2* and it was followed by marker assisted selection of HELP lines. Non-functional *Bna.FAD2* orthologues and functional *Bna.FAE1* orthologues resulted in the lower PUFAs and higher erucic acid in the HELP lines, respectively. Up to 63 % of very long chain fatty acids (VLCFAs, eicosenoic acid, C20:1 and erucic acid, C22:1) were reported in the HELP lines as compared to 55-56% in the control varieties. PUFAs levels were between 3-6 % in most of the lines as compared to 20-21% in the controls. The effect of knocking out *Bna.FAD2* family (reducing PUFAs) in the high erucic cultivars has demonstrated the increase of VLCFAs. Erucic acid is a valuable chemical for the industry and its high proportion in the oil will be useful for the industry.

P0597: Brassicas, Arabidopsis, and related

Clubroot Resistance Gene *Rcr3* Mapped in *Brassica rapa* using Bulk Segregant RNA Sequencing

Md Masud Karim, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada and Abdulsalam Dakouri, Zhang Yan, Qilin Chen, Gary Peng, Stephen E. Strelkov, Bruce D. Gossen, Fengqun Yu

Genetic resistance is widely used to manage clubroot (*Plasmodiophora brassicae*) in *Brassica* crops, but development of new pathotypes can result in rapid breakdown of resistance. In the current study, a new resistant loci (*Rcr3*) was mapped in *B. rapa* using bulk segregant RNA sequencing (BSR-seq). *Rcr3* was effective against the most prevalent pathotype on the Canadian prairies (pathotype 3) and a new pathotype (5X). BC_1 plants exhibited 1:1 segregation for resistance for both pathotypes, which indicated control by a single dominant resistance gene. In total 654.1K and 637.7 K variants were identified by assembling 107.5 M and 97.1 M reads against *B. rapa* "Chiifu" reference genome v1.5 from resistant and susceptible bulk. SNP variants were much more prevalent than Indels (89% vs. 10%). Chromosome A08 carried the highest number of polymorphic variants (38%) and Kompetitive Allele Specific PCR (KASP) analysis was used to select polymorphic SNPs from chromosome A08. *Rcr3* was mapped using 23 polymorphic SNPs with a range of 38.3 cM, where two flanking markers were selected 417 Kb apart. Analysis of differentially expressed genes identified six genes in the *Rcr3* region that were highly expressed in the R bulk, and gene annotation identified four genes associated with plant defence.

P0598: Brassicas, Arabidopsis, and related

Using *Arabis* Inter-Species Introgression Lines to Understand Divergence of Flowering Pathways in Annual and Perennial Plants

Eva Madrid-Herrero[1], Edouard Severing[1], Christiane Kiefer[2], Elisa de Ansorena[1], Wen-Biao Jiao[1], Korbinian Schneeberger[1] and George Coupland[1], (1)Max Planck Institute for Plant Breeding Research, Cologne, Germany, (2)Centre for Organismal Studies, Heidelberg, Germany

Little is known about the traits modified during evolution that lead to divergence of annual and perennial life cycles. In this study, we aim to identify genes that diverge in function that drive the emergence of life-history traits. To this end, a population of introgression lines (ILs) is being developed from an interspecific cross between the annual *Arabis montbretiana* and the perennial model *A. alpina*. Among these ILs, we identified an annual introgression on chromosome 2 that is significantly associated with a delay of flowering compared to the *A. alpina* parent. The effect of the introgression is dominant, suggesting a gain of function from the annual. Subsequent fine mapping reduced the region to a 0.7 Mb segment containing 150 genes. Of these genes a tandem-duplicated pair of genes, homologs of the *Arabidopsis thaliana MADS AFFECTING FLOWERING-3* (*MAF3*) gene are the most promising candidates for the flowering-time phenotype. The corresponding *A. alpina* genes accumulated several mutations and are unlikely to produce functional proteins. Phylogenetic analysis of these MADS box genes, in several *Arabis* genomes, point to a complex evolutionary history.

Preliminary expression studies showed that in the late-flowering ILs carrying the *MAF3-like* genes, the floral identity genes *AP1*, *FUL* and *LFY* are reduced in expression, while the floral activator *SOC1* is expressed similarly to the *A. alpina* parent. These results indicate that flowering could be delayed through a *FLC/SOC1* independent pathway. We are currently studying whether these genes have played an important role in life history divergence of the studied *Arabis* species

P0599: Cotton

CottonGen: An Online Resource for the Cotton Community

Jing Yu[1], Sook Jung[1], Chun-Huai Cheng[1], Taein Lee[1], Katheryn Buble[1], Ping Zheng[1], Jodi L. Humann[1], Deah McGaughey[1], Heidi Hough[1], Stephen P. Ficklin[1], B. Todd Campbell[2], Richard G. Percy[3], Don C. Jones[4] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)USDA-ARS, Florence, SC, (3)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (4)Cotton Incorporated, Cary, NC

CottonGen (www.cottongen.org) is a well curated, community-oriented informatics resource for cotton researchers that facilitates research discovery and cultivar improvement by curating, integrating, comparing, and maintaining a database that aims to serve as the central data repository for the cotton community. CottonGen not only contains genetic maps, QTLs, germplasm, markers, and genome sequences, but also has tools to view genomes (JBrowse and GBrowse), search DNA and protein sequences (BLAST+), view metabolic pathways (PathwayTools), and to view genetic maps (MapViewer and CMap). CottonGen now also uses a Cotton Trait Ontology for phenotype related data and has updated versions of the CottonGen Reference Transcriptomes (RefTrans), which are assemblies of peer-reviewed, published RNA-Seq and EST datasets for individual *Gossypium* species. CottonGen contains the whole genome sequences and annotations of three cultivated species (AD1, AD2, A2) and one wild species (D5), and the metabolic pathways of AD1 and D5. The CottonGen Breeders Information Management System (BIMS) is in development. BIMS is a convenient tool for management of cotton breeding programs and integrates both public and private breeding data for use with management tools. BIMS will also access data from several larger datasets such as the trial data from the Regional Breeders Testing Network (RBTN), the germplasm characterization data from the US National Cotton Germplasm Collection (NCGC), and the germplasm evaluations from the US The Germplasm Resources Information Network (GRIN), China, and Uzbekistan. CottonGen will continue to support the cotton research community by providing useful research tools and serve as a repository for data. CottonGen is directly supported by Cotton Incorporated, USDA-ARS, the cotton industry and USDA NRSP10.

P0600: Cotton

Agriplex PlexSeq™ as a Platform for High-Throughput SNP Screening of Cotton Germplasm

Robert Vaughn[1], David M. Stelly[1], Alex Miron[2], Brian Kevany[3] and Frank Campbell[3], (1)Texas A&M University, College Station, TX, (2)AgriPlex Genomics, Cleveland, OH, (3)Agriplex, Cleveland, OH

High-throughput screening of Cotton germplasm for interspecific SNPs facilitates rapid and efficient genomic introgression for trait discovery and crop improvement.

A limiting factor is cost: sequencing requires expensive equipment, reagents and high-quality DNA. PCR-based fluorometric assays are inexpensive and relatively tolerant of crude DNA preps. However, targeting multiple loci adds significantly to the cost and time. Ploidy-related issues present significant challenges to any genotyping platform, and PCR-inhibiting compounds in leaf and cotyledon tend to reduce amplification efficiency.

Agriplex Genomics' PlexSeq technology is a multiplex PCR platform for sequence-based SNP genotyping that attracted our attention as a prospective technical alternative to assay-based genotyping, and one that could be more affordable and robust than alternative sequencing-based approaches, especially for mid-ranged applications involving broad genome coverage.

DNA was extracted in-house from various genotypes of cotton and related species, using kit-based or crude DNA extraction methods, as well as by AgriPlex. Primers were developed from short (~200 bp) SNP-containing sequences that were roughly evenly spread along the genome for ~10 per chromosome (n=26) for a total of ~260.

Using kit-extracted DNA, 96.8% of data points were readable. For crude extracts, a smaller subset was used (only 17 primers and lower numbers of samples), but the success rate was 99.5%. Crude cotyledon extracts yielded a 97.6% success rate, but crude leaf extract was largely unsuccessful (85% bad calls).

P0601: Cotton

Single-Molecule Sequencing of *G. arboreum* Genome and GWAS Analysis of Chinese *G. arboreum* Accessions

Xiongming Du, Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang,Henan, China

The allotetraploids of cotton plants are designated as AADD genomes. The diploid ancestral A of the allotetraploid are thought to have diverged from *G. arboreum* or *G. herbaceum*. Here, a higher quality *G. arboreum* genome assembly was got by using PacBio single molecule real-time sequencing technology sequencing and High-throughput chromosome conformation capture (Hi-C) technology. We generated 142.54 Gb of raw PacBio reads (approximately 77.6-fold genome coverage) assembled 8,223 contigs with a N50 length of 1.1 megabases. 40,960 genes were annotated assistant by the PacBio transcriptome analysis. We further sequenced 230 *G. arboreum* accessions (average 6× depth) to generate a map of genome variation including~18 million SNPs and ~2 million indels. A phylogenetic tree based on SNP genotypes, together with an analysis of population structure suggested that Chinese *G. arboreum* originated in South China and was subsequently introduced to the Yangtze River and Yellow River regions. A total of 98 significant peak associations (-logP>6) for 19 agronomically-important traits in *G. arboretum* were identified in our genome-wide association study (GWAS). Most of the population divergence related traits, including yield and disease resistance, have experienced geographic isolation. Gain of *Fusarium wilt* disease resistance in YZ and YR accessions is associated with *GaGSTF9*, whose expression is highly inducible upon fungi inoculation. A non-synonymous SNP mutant substitution of *GaKASIII* seems to have conferred significant alteration of the fatty acid composition (C16:0 and C16:1) in cotton seed. We further identified a new QTL related to seed fuzz development by combining our GWAS analysis with bulk population sequencing. Our work provides abundant resources for cotton genetics study and molecular breeding, and deepens our knowledge about A-genome evolution, especially Chinese *G. arboreum* population.

P0602: Cotton

Domestication and Improvement of Cotton: From Tree Cotton to American Upland Cotton and to the World's Largest Fiber Crop

Lei Fang[1], Yan Hu[2], Qiong Wang[3], Jiedan Chen[1], Xueying Guan[1,3], Hao Gong[4], Yinhua Jia[5], Chunxiao Liu[3], Baoliang Zhou[3], Wangzhen Guo[3], Xuehui Huang[6], Xiongming Du[5] and Tianzhen Zhang[2], (1)Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China, (2)Zhejiang University, Hangzhou, China, (3)Nanjing Agricultural University, Nanjing, China, (4)National Center for Gene Research, Chinese Academy of Sciences, Shanghai, China, (5)Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Ministry of Agriculture, Anyang, China, (6)Shanghai Normal University, Shanghai, China

Allotetraploid upland cotton (*Gossypium hirsutum* L.) provides the most important natural fiber in the world. It was initially domesticated from tree cotton in the Mesoamerican and Caribbean regions, and subsequently further domesticated and improved in the southern United States, from the perennial to the annual crop. Because of its high-yield property, American upland cotton was extensively introduced and grown in Asian countries such as India and China. To unfold this evolution and domestication history, we resequenced the genomes of 147 cotton accessions, including diverse wild relatives, landraces, and modern cultivars. By comparing the genetic diversity among wild *G. hirsutum* cultivars and races, we identified 109 domestication-related selective sweeps, including 723 fiber related and 115 seed germination-related genes, indicating the contribution of selective sweeps in the domestication for fiber yield and fiber qualities. To further unveil the improvement history, we report a comprehensive genomic assessment of modern improved upland cotton based on the genome-wide resequencing of 318 landraces and modern improved cultivars. More associated GWAS loci for lint yield (71) are detected than those for fiber quality (45), which suggests that lint yield has stronger selection signatures than other traits. Moreover, we evaluated the contributions of various genetics pools to the current cultivars and found that 54.8% of the elite GWAS alleles detected were transferred from three founder landraces. Our findings uncover the domestication and improvement history of allotetraploid upland cottons and provide genomic bases for improving cotton product and for further evolution analysis of polyploid crops.

P0603: Cotton

The Role of Noncoding RNA Pattern and Function in Cotton Fiber Cell Development

Xueying Guan, Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China

Cotton fiber is the most important sustainable fiber source for textile industry. It is a single cell organ derived from the epidermis of the cotton ovule or seed. To understand the molecular basis of plant cell differentiation pattern, the cotton fiber cell is a good model system. Non-coding RNA is emerging as one of the most important regulators for the gene expression in response to multiple biological transitions and environmental stimuli. We systematically investigate the non-coding RNA behavior in the fiber cell differentiation progress. The major data indicate both small RNA and long non-coding RNA (lncRNA) play critical roles in fiber cell development. First of all, the fiber cell generates a unique group of small RNA in the fiber initiation stage. For example, the miR828 and miR858 trigger the target gene GhMYB2 to generate tasiRNAs in fiber cell fate determination. Another MIXTA MYB transcriptional factor coding gene, GhMML3 can generate an antisense transcript on the 3' end of the gene loci. Together with the sense and antisense transcripts, a double strand RNA come into being to derive small RNAs. These secondary generated small RNAs interfere the cell fate determination in the mml3 mutant in stimulating the fiberless seed phenotype. On the other hand, the long non-coding RNAs are also found to take parts in the fiber cell differentiation by small RNA generation. We therefore conclude noncoding RNAs directly impact the fiber cell development in multiple aspects of molecular regulation.

P0604: Cotton

A Perspective of Genomic Selection Application in Upland Cotton Breeding

Md Sariful Islam, Sugarcane Production Research Unit, USDA-ARS, Canal Point, FL, Salvador Gezan, University of Florida, Gainesville, FL, Johnie N Jenkins, USDA-ARS, Mississippi State, MS, Chris D. Delhom, USDA ARS SRRC, New Orleans, LA, Jack McCarty, Genetics & Sustainable Agriculture Research Unit, USDA-ARS, Mississippi State, MS and David D. Fang, Cotton Fiber Biosciences Research Unit, USDA-ARS-SRRC, New Orleans, LA

The use of genomic selection (GS) has stimulated new ways to utilize molecular markers in breeding for complex traits in absence of phenotypic data. Numerous statistical models and approaches have been proposed in several crops; however, no comparative analysis is available, at the present, to identify the most promising ones in Upland cotton. The objective of this study was to experimentally evaluate the potential of genomic selection in Upland Cotton breeding. Six fiber quality traits data obtained from three years replicated field trials in Starkville, MS. Genotyping-by-Sequencing (GBS) based genotyping was performed using 547 recombinant inbred lines (RILs) of the multi-parent advanced generation inter-cross (MAGIC) population, and 6,071 Single Nucleotide Polymorphic (SNP) markers were selected for the GS analysis. In addition, 224 microsatellite genotype data were also incorporated in this GS analysis. Several methods were compared, including Genomic BLUP (GBLUP), Ridge Regression BLUP, Bayes B, Bayesian LASSO Regression, and Reproducing Kernel Hilbert Spaces. The average heritability (H^2) ranged from 0.56 to 0.83 for all tested traits across the three years evaluated. The prediction ability (PA) and prediction accuracy (PACC) for the genomic estimated breeding value using GBLUP varied widely (0.16 to 0.54 and 0.28 to 0.57, respectively) across the three years for all tested traits with the highest PA and PACC at 2010 for fiber elongation. Results indicated that GS could predict GEBV efficiently in Upland cotton fiber quality attributes and has great potential utility in breeding by reducing cost and breeding cycle.

P0605: Cotton

Elucidating the Roles of microRNAs in Cotton Fiber Initiation and Early Development

Qinglian Wang, Henan Institute of Science and Technology, Xinxiang, China and **Baohong Zhang**, East Carolina University, Greenville, NC

Cotton fiber development is a fundamental biological process; investigating cotton fiber initiation and development provides a unique window into the regulation of cell differentiation, cellulose biosynthesis, and further increasing cotton fiber quality and yield. For several decades, a great deal of research has been aimed at elucidating the underlying molecular pathways. Nevertheless, the mechanisms by which cotton fiber differentiates and develops remains unclear. In this study, we employed high throughput deep sequence, degradome sequence as well as quantitative real time PCR to identify and functionally analyze microRNAs (miRNAs), an important gene regulator, in both fiberless mutants as well as its wildtypes. We found that both conserved and novel miRNAs have unique expression pattern in cotton fiber development. During the cotton fiber development, particularly at the 10 DPA, lots of miRNAs show different modification that suggests miRNA-regulating cotton fiber through miENA modification. The miRNA genes controlling cotton fiber development are majorly from subgenome A. Using transgenic, RNAi and genome editing technologies show that overexpression and knockout/knockdown of an individual miRNAs affected cotton fiber development and further affect cotton fiber length and quality.

P0606: Cotton

The Role of Noncoding RNA Pattern and Function in Fiber Cell Development

Xueying Guan, Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China and Ting Zhao, Nanjing Agricultural University, China

Cotton fiber is the most important sustainable fiber source for textile industry. It is a single cell organ derived from the epidermis of the cotton ovule or seed. To understand the molecular basis of plant cell differentiation pattern, the cotton fiber cell is a good model system. Non-coding RNA is emerging as one of the most important regulators for the gene expression in response to multiple biological transitions and environmental stimuli. We systematically investigate the non-coding RNA behavior in the fiber cell differentiation progress. The major data indicate both small RNA and long non-coding RNA (lncRNA) play critical roles in fiber cell development. First of all, the fiber cell generates a unique group of small RNA in the fiber initiation stage. For example, the miR828 and miR858 trigger the target gene *GhMYB2* to generate tasiRNAs in fiber cell fate determination. Another MIXTA MYB transcriptional factor coding gene, *GhMML3* can generate an antisense transcript on the 3' end of the gene loci. Together with the sense and antisense transcripts, a double strand RNA come into being to derive small RNAs. These secondary generated small RNAs interfere the cell fate determination in the *mml3* mutant in stimulating the fiberless seed phenotype. On the other hand, the long non-coding RNAs are also found to take parts in the fiber cell differentiation by small RNA generation. We therefore conclude noncoding RNAs directly impact the fiber cell development in multiple aspects of molecular regulation.

P0607: Cotton

Genome-Wide Identification and Characterization of SPL Transcription Factor Family and their Evolution and Expression Profiling Analysis in Cotton

Caiping Cai[1], Wangzhen Guo[1] and **Baohong Zhang**[2], (1)Nanjing Agricultural University, Nanjing, China, (2)East Carolina University, Greenville, NC

Plant specific transcription factors, SQUAMOSA promoter-binding protein-like (SPL), are involved in many biological processes. However, no systematical study has been reported in cotton. In this study, a total of 177 *SPL* genes were identified, including 29, 30, 59 and 59 *SPLs* in *Gossypium arboreum*, *G. raimondii*, *G. barbadense,* and *G. hirsutum*, respectively. These *SPL* genes were classified into eight phylogenetical groups. The gene structure, conserved motif, and clustering were highly conserved within each orthologs. Two zinc finger-like structures (Cys3His and Cys2HisCys) and NLS segments were existed in all GrSPLs. Segmental duplications play important roles in SPL family expansion, with 20 genes involved in segmental duplications and 2 in tandem duplications, and ten ortholog pairs in syntenic regions between *G. raimondii* and *A. thaliana*. Several putative cis-elements, involved in light, stresses and phytohormones response, were found in the promoter regions of *GhSPLs*, suggesting that plant responses to those environmental changes may be induced through targeting SPL transcription factors. RNA-seq analysis shows that SPL genes were differentially expressed in cotton; some were highly expressed during fiber initiation and early development. Comparing with other plants, SPL genes show subfunctionalization, lost and/or gain functions in cotton during long-term domestication and evolution.

P0608: Cotton

Identification and Expression Analysis of Phosphatidy Ethanolamine-Binding Protein (PEBP) Gene Family in *Gossypium hirsutum* L.

Baohong Zhang[1], Yangguang Tan[2] and Qinglian Wang[2], (1)East Carolina University, Greenville, NC, (2)Henan Institute of Science and Technology, Xinxiang, China

The phosphatidy ethanolamine-binding protein (*PEBP*) family genes play an important role in generating mobile flowering signals and controlling floral development and timing. Here, we identified a total of 40 *PEBP* genes, in which 20, 10, and 10 were from three cotton species: tetraploid Gossypium hirsutum. (AD1), diploid *G. raimondii* (D5) and *G. arboreum* (A2), respectively. In *G. hirsutum*, the 20 identified *PEBP* genes were unevenly distributed on 12 chromosomes except four were located on the scaffolds. The identified *PEBP* genes were classified into four groups (*TFL1, MFT, FT and FT-like*); the majority of *PEBP* genes had similar intron/exon distribution, whereas the divergence of *PEBP* genes suggests the possibility of functional diversification. The expression of *PEBP* genes varied among different tissues. The expression of *GhFT1D* and *GhFTL2D* were significantly higher in leaves, *GhFTL1A, GhFTL1D* and *GhFT1A* has a highest expression level in fiber, whereas *GhFTL2A*, *GhMFT3A* and *GhMFT3D* had higher expression level in petal or stamen. This study brings new insights into the integrated genome-wide identification of *PEBP* genes in cotton and provides references for promoting early maturation in cotton breeding.

P0609: Cotton

Toward Development of Cotton with Glandless Seeds and Glanded Plants

Yan Hu[1], Dan Ma[2], Lei Fang[1] and Tianzhen Zhang[3], (1)Zhejiang University, College of Agriculture and Biotechnology, Hangzhou, China, (2)Institute of cotton research of CAAS, Anyang, China, (3)Zhejiang University, Hangzhou, China

Besides cotton fibers, cottonseeds are the main product of cotton plant. However, gossypol deposited in the glands makes cottonseed toxic to non-ruminants, hampering the commercialization of cottonseed protein and oil for human consumption but important resistant substance of *Gossypium* and storages in pigment gland. Previous researches have shown gossypol content was positive relevant to number, size and density of glands. With or without glands, their densities and sizes are often used as an indicator for high or low gossypol content in cotton. Through previous study, we identified the dominant glandless gene Gl_2^e by map-based cloning, which is proved to be a key factor in gland formation and named *GoPGF* (Gossypium pigment gland forming). Knock-down expression of Gl_2^e in the glanded cotton generated glandless phenotype and strongly decreasing gossypol content in the leaves which was consistent with the dominantly suppressed transcript levels of genes involved in the terpenoid biosynthesis pathways, as well as TPSs. This finding provides an important theoretical basis for effective regulation of gossypol at a molecular level and also new avenues for producing produce 'glandless seed and glanded plant' by genetic engineering.

Keywords: cotton, pigment gland, gossypol, map-based cloning

P0610: Cotton

Regulation of Vertical Na+ Transport and Partitioning and its Contributions to Natural Variation for Salt Tolerance across the Cultivated Tetraploid Gossypium Germplasm

Kevin R. Cushman[1], Lori L. Hinze[2], Isaiah Pabuayon[1], Megan Sweeney[3] and Benildo G. de los Reyes[1], (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (3)Bayer Crop Science Division, Morrisville, NC

Physio-morphometric evaluation of a subset of accessions representing the core of allelic variation across the *Gossypium* Diversity Reference Set (GDRS) at low to moderate salt stress suggested a widely occurring avoidance mechanism that makes *Gossypium* more salt-tolerant than many other crops. We hypothesized that this mechanism is conferred by delayed onset of injury by slowing the progression of Na^+ accumulation from source to the more sensitive sink tissues. We addressed this hypothesis by examining the vertical expression profiles of major transporter genes involved in intracellular and intercellular Na^+ partitioning from the roots to progressively younger leaves and shoot as sinks. We observed significant changes in vertical expression of *ghNHX1*, *ghSOS1*, and *ghHKT1* across source and sink tissues with increasing severity of salt stress. Levels of cellular Na^+ and the direction of Na^+ export varied across genotypes. Current results suggest that extrusion of Na^+ across the plasma membrane via *ghSOS1* occurs most at mild stress, mainly in older leaves. Na^+ transport across the tonoplast via *ghNHX1* was most prominent at higher stress levels and most apparent in the roots and younger sink tissues such as the shoot. The most tolerant cultivars are better in avoiding the effects of salt stress through vacuolar sequestration mediated by increased expression of *ghNHX1*. Genotypic differences in vertical Na^+ partitioning appeared to correlate with biomass accumulation, overall plant health and vigor based on standard evaluation scores (SES), increased antioxidant activity in the shoots, and reduced proline levels and cell membrane destabilization.

P0611: Cotton

Transcriptome Analyses of Reniform Nematode Infested and Uninfested Susceptible and Tolerant Genotypes of Upland Cotton (*Gossypium hirsutum*)

Kadiatou Keita[1], Abreeotta Williams[2], Matthew Miller[1], Seloame T. Nyaku[3], Kathy Lawrence[4], Govind C. Sharma[1] and Venkateswara R. Sripathi[1], (1)Alabama A&M University, Normal, AL, (2)Columbia State Community College, Columbia, TN, (3)University of Ghana, Accra, Ghana, (4)Auburn University, Auburn, AL

Reniform nematode (*Rotylenchulus reniformis*) is one of the world's economically perilous phytoparasitic nematodes. It is pathogenic to over 300 crop species while cotton is the preferred host. Reniform nematode (RN) is currently considered as a significant threat to the world food supply when combined with other bacterial and fungal diseases. This study aimed at generating resources to fight against RN infestation by screening the transcriptomes of susceptible and tolerant genotypes of *Gossypium hirsutum*. Two genotypes of upland cotton (LONREN and FiberMax) were germinated in 0.5 L pots with 3/4th of sterilized soil (60% sand and 40% silt + clay). The plants were exposed to different concentrations of nematodes (0, 5k and 50k) twenty-days after germination. The leaf and root samples were collected from the plants 21-days after inoculation, and then RNA was isolated for RNA-Seq analyses. The experiment included two genotypes (LONREN and FiberMax), three treatment conditions (0, 5k and 50k) and three replicates (R1, R2, and R3) to generate 18 RNA-Seq libraries and sequenced on Illumina Hiseq2500 to generate over 500 million paired-end reads (2 x 100 bp). The transcriptomic analyses showed that 28,658 genes in LONREN and 30,245 genes in FiberMax were differentially expressed when compared to the control plants. Among these, 17 significantly overexpressed genes in LONREN were identified when compared to FiberMax, which has a crucial role in the cell wall and secondary metabolite synthesis. Also, two classes of R-genes that have established roles in other biotic stresses were identified in LONREN. Furthermore, this study serves as a prelude to the integration of mRNA-Seq, smRNA-Seq, and degradome sequencing data from LONREN and FiberMax in identifying relevant genes and pathways associated with resistance and susceptibility of RN in cotton.

P0612: Forest Trees

TreeGenes: Turning over a New Leaf

Emily Grau[1], Sean Buehler[1], Nic Herndon[1], Taylor Falk[1], Peter Richter[1], Margaret Staton[2], Stephen P. Ficklin[3] and Jill L. Wegrzyn[1], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Tennessee, Knoxville, Knoxville, TN, (3)Washington State University, Pullman, WA

TreeGenes is a web-based information resource designed to serve the diverse needs of the forest tree genomics research community by uniting information resources with tools visualization and analysis tools. TreeGenes has recently undergone a complete redesign using Tripal, a tool to create and manage genomic database websites. An open source project, Tripal allows developers the flexibility to create their own tools (modules) as well as open communication among Tripal supported repositories.

TreeGenes hosts a range of modules that have expanded functionality to deliver genomic and phenotypic data on >1700 forest tree species. The new Galaxy module allows users to execute analytical workflows with next generation sequencing datasets on high performance computing resources with the click of a button. The Elasticsearch module allows flexible searching and data retrieval between sites such that TreeGenes can share data with Hardwood Genomics Project and the Genome Database for Roseaceae.

TreeGenes is also creating new modules, including: CartograTree which integrates envrionmental, phenotypic, and genotypic data for georeferenced trees. The module works with the Tripal Galaxy to facilitate association mapping and landscape genetics analysis. Source data for CartograTree is imported via the Tripal Plant PopGen Submit module, a pipeline for accepting direct submissions from researchers. This module collects relevant metadata while reducing the burden on the researcher for submission. The new TreeGenes Tripal DIAMOND module offers speed improvements over BLAST and allows sequence similarity searches across numerous genomes and transcriptomes. TreeGenes looks forward to providing the forest tree research community with expanded data and analytical toolsets.

P0613: Forest Trees

OrthoQuery - a Tripal Module for Supporting Identification and Visualization of Orthogroups

Sumaira Zaman[1], Emily Grau[2], Sean Buehler[2], Stephen P. Ficklin[3] and Jill Wegrzyn[2], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of Connecticut, Storrs, CT, (3)Washington State University, Pullman, WA

Despite the decreasing cost of sequencing, significant computational challenges exist in assembling a well resolved genome. Hence, many scientific communities are studying transcriptomes to generate gene models and answer questions in comparative genomics. Orthologous genes, or orthogroups, have proven to be an important tool for comparative genomics even in the absence of a reference genome. Comparisons across orthogroups can help characterize selection pressure, evolutionary rate of specific gene families, novel gene families, and identify whole or partial genome duplication events. Tools like OrthoFinder and OrthoMCL can execute comparisons and investigate relationships among large sets of protein sequences. These tools are however limited as they do not allow users to select and filter datasets, or provide a robust visualization and query interface. We propose Orthoquery, a Tripal module that will provide a semi-automated analytical pipeline and visualization platform. The OrthoQuery Tripal Module will identify orthologous genes from forest tree species represented in the TreeGenes database, Orthofinder optimized with Diamond as the tool for protein level comparisons, and Tripal framework with Galaxy integration to support visualization and query. Although development of this module is intended for forest tree research, other clade or organism specific data providers will have access to this pipeline since it will be provided to Tripal's open source project. The customizable module will have the ability to integrate with any of the other 30+ Tripal supported databases. With the development of our pipeline, we hope to aid other researchers in uncovering various phylogenetics relationships that otherwise would go undetected.

P0614: Forest Trees

Efficiency, Design, and Analytical Improvements in Cartogratree, a Web-Framework for Association Mapping and Landscape Genomics in Forest Trees

Taylor Falk, Nic Herndon, Emily Grau, Sean Buehler, Peter Richter and Jill L. Wegrzyn, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT
CartograTree is a web-based framework developed, in conjunction with the TreeGenes project, to create an efficient tool for researchers to display, select, analyze, and document model and non-model trees in conjunction with their associated genotypic and phenotypic metrics. The integration of environmental layers with genomic data associated with these georeferenced trees allows for analyses such as association mapping and landscape genomics. Significant updates are implemented to increase the usability of CartograTree, including an advanced query interface, plans for updated plotting and selection tools, references to publications sourced from TreeGenes and Dryad repositories, as well as a wide range of environmental and climatic variables. CartograTree is also integrated into the new Tripal Plant PopGen Submit (TPPS) module, which allows detailed collection of data and metadata associated with publications. Since TreeGenes is running on the Tripal 3.0 platform, the recent integration of the web-based next general sequencing toolkit, Galaxy, is integrated to provide direct access to analytical pipelines. In CartograTree, this integration enables cross-site querying with partner tree databases, and analysis focused on association mapping. CartograTree currently features a subset of the approximately 1,700 forest tree species found in TreeGenes, representing over 50 unique species and more than 48,000 individual tree accessions from TPPS and Dryad publications. The responsive and visually appealing CartograTree framework will help users perform efficient and comprehensive genomics analysis that can be applied to other plant systems.

P0615: Forest Trees

Purifying Selection Patterns in Genes of Two Distantly Related Conifers

Mengmeng Lu[1], Kathryn A. Hodgins[2], Sally N. Aitken[3] and Sam Yeaman[1], (1)University of Calgary, Calgary, AB, Canada, (2)Monash University, Melbourne, Australia, (3)University of British Columbia, Vancouver, BC, Canada
Conifer species inhabiting heterogeneous environments demonstrate phenotypic and genomic signatures consistent with local adaptation. Previous studies have identified genes with variants associated with spatial variation in temperature or cold hardiness in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca*, *Picea engelmannii*, and their hybrids), providing evidence of convergent local adaptation. However, if the intensity of purifying selection varies with the environment, clines in nucleotide diversity could evolve that would yield allele frequency-environment signatures resembling local adaptation. If similar patterns in the strength of purifying selection operate in different species, it may result in the appearance of convergent local adaptation. To test this possibility, we analyzed 86 natural populations of lodgepole pine and 77 of interior spruce (n = 3 trees/population) spanning heterogeneous environments in British Columbia and Alberta. We calculated nucleotide diversity of genes in each population and correlation between nucleotide diversity and five environmental variables used in previous analyses of local adaptation. Overall, we found no similar pattern in the convergent genes between the two species. When comparing the strength of correlations among genes with or without convergent adaption signatures, no disparity was shown in lodgepole pine. On the contrary, significant differences were detected in interior spruce, which may arise due to a combination of selection and hybridization. In conclusion, the results rule out the possibility that spatially variable purifying selection underlie the convergent local adaption signatures within the two species.

P0616: Forest Trees

Tissue-Specific Transcriptome Analyses of Wood Formation in Both Poplar and Pine Tree for Evo-Devo Study

Min-Ha Kim[1], Eung-Jun PARK[2], Hyoshin Lee[2] and **Jae-Heung Ko**[1], (1)Kyung Hee University, Yongin, South Korea, (2)National Institute of Forest Science, Republic of Korea, Suwon, South Korea
Woody biomass is one of the most abundant, eco-friendly and renewable source of alternate energy in near future. Wood (i.e., secondary xylem) is formed in the process called wood formation, which includes the vascular cambium differentiation and xylem cell fate specification. Vascular cambium is the cylindrical secondary meristem resides in stems and roots of perennial woody species. However, the underlying molecular mechanism of wood formation is still largely unknown. To get an insight on the wood formation through 'Evo-Devo' genomics approach, we performed a series of tissue/cell type-specific transcriptome analysis from both poplar and pine tree, which represent angiosperm and gymnosperm respectively, by using Next-Generation Sequencing (NGS) based RNA-Sequencing technology. Each cambium and xylem cells were isolated from vertical stem segments that represent a gradient of developmental stages regarding wood formation (i.e., immature, intermediate and mature stem). For references, shoot apical meristem (SAM) and leaf tissues were included. Firstly, quality of our sampling strategy was assessed by checking the cell-type specific expressions of several marker genes involved in cambium maintenance and secondary wall formation using qRT-PCR. Next, RNA-Seq was performed and obtained the cell type-specific transcriptome data. Further analysis identified a number of novel genes possibly involved in the cambium initiation/maintenance or xylem cell fate specification. Thus, our efforts will serve as a springboard for future studies aimed at unraveling molecular mechanism of wood formation, which is one of the most important biological processes on Earth. This work was funded by the Forest Resources Genome Project (2014071G10-1722-AA04).

P0617: Forest Trees

Designing a Genotyping Array for Genomic Selection in Loblolly Pine

Madison Caballero[1], Fikret Isik[2], Juan José Acosta[3], Trevor Walker[2], Richard Sniezko[4], Andrew J. Eckert[5] and Jill Wegrzyn[1], (1)University of Connecticut, Storrs, CT, (2)North Carolina State University, Raleigh, NC, (3)NC State University, Raleigh, NC, (4)USFS Dorena Genetic Resource Center, Portland, OR, (5)Department of Biology, Virginia Commonwealth University, Richmond, VA
Loblolly pine (*Pinus taeda*) is extensively cultivated for timber and pulp production in the Southeastern United States. It represents one of the largest genomes assembled to date at 22Gbp. To apply genomic selection to wood quality, disease resistance, and related production traits in large plantation populations, extensive genotyping is necessary. Given the large and fragmented genome, genotyping arrays provide a more cost effective and accurate platform to generate extensive SNP resources. This USDA-funded PineSNP project is leveraging resources (RAD-Seq and exome capture) of over 2,051 individuals from the previous PineMAP project. For the identification of variants, short reads from the exomes and GBS studies were aligned to the 2.01 version of the genome and bi-allelic SNPs were identified. Probes were designed from these variants considering overall quality, depth of coverage, polymorphic flanking sequence, and conversion success estimates as provided by

Illumina's ADT scoring. Annotations of the high-quality SNPs were assigned using a custom SNPeff database and the latest v2.01 annotation. The extensive filtering pipeline reduced the initial number of detected variants from 1,713,000 SNPs to 54,722 high-quality probes. These SNP annotations will be further evaluated for population metrics before inclusion in a final array that can be applied to genomic selection goals in loblolly pine.

P0618: Forest Trees

Population Genomics Supports Speciation with Gene Flow, Not Genomic Islands of Differentiation, in Sky-Island Populations of Southwestern White Pine

Justin C Bagley[1,2], Mitra Menon[1], Christopher Friedline[1], Amy Whipple[3], Anna Schoettle[4], Alejandro Sáenz[5], Christian Wehenkel[5], Lluvia Flores-Renteria[6], Richard Sniezko[7], Sam Cushman[8], Kristen Waring[9] and Andrew J. Eckert[1], (1)Department of Biology, Virginia Commonwealth University, Richmond, VA, (2)Departamento de Zoologia, Universidade de Brasília, Brasília, Brazil, (3)Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, (4)USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO, (5)Instituto de Silvicultura e Industria de la Madera, Universidad Juárez del Estado de Durango, Durango, Mexico, (6)Department of Biology, San Diego State University, San Diego, CA, (7)USFS Dorena Genetic Resource Center, Portland, OR, (8)USDA Forest Service, Rocky Mountain Research Station, Flagstaff, AZ, (9)School of Forestry, Northern Arizona University, Flagstaff, AZ

Understanding speciation, including processes leading to lineage divergence and the origin and maintenance of reproductive barriers, is a fundamental goal of evolutionary biology. During ecological speciation, populations initially diverge through disruptive selection forming different ecotypes that subsequently develop reproductive isolation, forming ecologically differentiated species. Under two models proposed to explain the maintenance of species boundaries during ecological speciation—the 'tension zone' and 'bounded hybrid superiority' models, we expect varying demographic scenarios with different genomic signatures, especially patterns of gene flow. Genetic tests of these models based on broad sampling across the ranges of hybridizing species can shed light on the importance of intrinsic versus extrinsic factors in maintaining species boundaries. Here, we use demographic modeling with genome-wide SNP data to test predictions on the prevalence of gene flow during species formation in two species of North American pine trees, *Pinus strobiformis* (southwestern white pine) and *P. flexilis* (limber pine), that are broadly distributed across the desert southwest. Our results strongly support a pattern of *P. strobiformis–P. flexilis* speciation with gene flow, as well as low–moderate ongoing gene flow, but fail to support models with islands of divergence (parameterized with heterogeneous migration). These findings broadly agree with predictions of the bounded hybrid superiority model and, along with other forthcoming results of the project, implicate a greater relevance of extrinsic factors in facilitating speciation in conifers.

P0619: Forest Trees

Two Rounds of Ancient RNA-Mediated Gene Duplication Revealed by Whole Genome Mining of TE-Enriched Conifer Genomes

Jian-Feng Mao, Beijing Forestry University, Beijing, China

Gene duplication is an important mechanism for new gene origination and adaptation, in turn driving organismal evolution. RNA-mediated gene duplication or gene retroposition, the formation of gene retrocopies via the enzymatic machinery encoded by retrotransposable elements, was observed as key evolutionary process for developmental innovation and adaptive transition in primarily mammals, fruit flies and model plant. However, knowledge on expansion and evolutionary dynamics of RNA-mediated gene duplication is still limited, especially for the TE-enriched non-model genomes (like conifer genomes) for which retrotransposable elements are prevalent and TE removal is inefficient. By whole genome examining, we detected tens thousands of intact, pre-stop and frame-shit conveying retrocopies for each of the four released conifer (*Pinus taeda*, *Pinus lambertiana*, *Picea abies*, *Pseudotsuga menziesii*) genomes and Ginkgo genome. Two ancient rounds of retrocopy expansion were recovered by plotting Ks distribution, with the most ancient round associates with the origination of land plant, the other with separation between angiosperm and gymnosperm. Further survey in TE-enriched seed plant genomes like wheat, maize and cotton support all two expansion events revealed here, fern and moss genomes support the most ancient one, but no ancient expansion detected in ancestry lineage of land plant, like alga. Tissue specific expression of the retrocopies was also analyzed in conifer genomes. Our study provide valuable information on understanding the complexity of conifer genome and gene retroposition in plant.

P0620: Forest Trees

Transcriptome Analysis during Adventitious Root Formation in *Cryptomeria japonica*

Yuki Fukuda[1,2], Tomonori Hirao[3], Kentaro Mishima[1], Mineko Ohira[1], Yuichiro Hiraoka[1], Makoto Takahashi[1] and Atsushi Watanabe[4], (1)Forest Tree Breeding Center, FFPRI, Hitachi, Ibaraki, Japan, (2)2Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Fukuoka, Japan, (3)Forest Bio-Research Center, FFPRI, Hitachi, Ibaraki, Japan, (4)Faculty of Agriculture, Kyushu University, Fukuoka, Japan

Adventitious root formation of cuttings is an essential physiological process for the successful cutting propagation which maintains the genotype of the donor plant, and hence the superior characteristics of targeted clones. The ability of adventitious root formation is one of very important traits in breeding of various horticultural and forest tree species. Though understanding of biological processes which is a very complex and influenced by various factors is of considerable significance, little is known about molecular events such as changes in gene expression profiles that occur during adventitious root formation in forest trees. In this study, we attempted to characterize transcriptome of adventitious root formation in *Cryptomeria japonica* with microarray analysis. Firstly, we revealed temporal alteration of global gene expression, suggesting that major turning points of gene expression in adventitious root formation occurred at a faster time point before morphological changes. Secondary, we clarified the expression pattern of the candidate genes related to metabolism of carbohydrate or plant hormones, the molecules suggested to be associated with adventitious root formation, and generally supported the biochemical and molecular biological findings in other plants. Our results will provide the molecular biological basis to understand the physiological mechanism underlying adventitious root formation in coniferous species such as *C. japonica*.

P0621: Forest Trees

Characterization of the Molecular and Cellular Role of EVE1 in Plant Vascular Development

Johnathon Blahut, Kelly Balmant, Daniel Conde, Christopher Dervinis, David Oppenheimer and Matias Kirst, University of Florida, Gainesville, FL

A remaining hallmark of the colonization of terrestrial environments is the variation in complexity of the water-conducting tissues across the plant kingdom. The vascular cells of mosses and liverworts can be surprisingly complex with thickened cell walls and plasmodesma-derived perforations. These features mimic the water-conducting tissues observed among the angiosperm clade which has adopted vessels for improved conductive efficiency. In most cases, vessels show greater hydraulic conductivity owing to intervessel pits and perforation plates. To better understand this complexity, we mapped quantitative trait loci associated with vessel-related traits and discovered the *Enlarged Vessel Elements 1* (*EVE1*) gene in poplar. Stems of hybrid poplar trees overexpressing *EVE1* display higher hydraulic conductivity, vessel number per sapwood area and vessel diameter without a concomitant increase in cavitation. Further analysis of relative transcript abundance among unrelated individuals suggests a role of *EVE1* in determining vessel size. Recent protein electrophoresis studies have revealed that the *EVE1* locus produces a small molecular weight membrane protein in *Populus deltoides* stems. Immunolocalization performed with sapling stem sections suggest that the EVE1 protein localizes specifically to developing vessel elements but is largely absent of mature vessels. The discovery and characterization of EVE1 represents a starting point towards identifying other genes involved in plant vascular transport.

P0622: Forest Trees

Full Length Isoform Sequencing and Profiling of Expressed Genes in Cambial Zone and Needle Leaf of Japanese Larch (*Larix kaempferi*)

Kentaro Mishima[1], Hideki Hirakawa[2], Taiichi Iki[3], Tomonori Hirao[4], Akira Tamura[5], Manabu Kurita[6] and Makoto Takahashi[1], (1)Forest Tree Breeding Center, FFPRI, Hitachi, Ibaraki, Japan, (2)Tohoku University, Kisarazu, Japan, (3)Tohoku Regional Breeding Office, Forest Tree Breeding Center, FFPRI, Takizawa, Iwate, Japan, (4)Forest Bio-Research Center, FFPRI, Hitachi, Ibaraki, Japan, (5)Forest Breeding Center, Forestry and Forest Products Institute, Hitachi, Japan, (6)Kyushu Regional Breeding Office, Forest Tree Breeding Center, FFPRI, Koshi, Kumamoto, Japan

Japanese larch (*Larix kaempferi*) is a deciduous coniferous species that relies on wind-mediated pollen and seed dispersal, and it is one of the most important forestry tree species in northern Japan.The objective of this study is to produce an extensive collection of sequenced full length isoform (ESTs) found in xylem, needle leaf in Japanese larch. The collected isoforms are to be reference sequence for using various experimental approach. For this purpose, we identified 79,832 isoforms by PacBio RSII using isoform sequencing library from the cambial zone and needle leaf in Japanese larch. To gain insight into seasonal expression patterns cambial zone and needle leaf, a custom cDNA microarray was designed from the isoforms obtained and investigated differential gene expression in Japanese larch.

P0623: Forest Trees

Identification of QTL for Pine Wood Nematode Resistance in Japanese Black Pine (*Pinus thunbergii*) Using Genotyping-By-Sequencing (GBS)

Tomonori Hirao, Forest Bio-Research Center, FFPRI, Hitachi, Ibaraki, Japan and Koji Matsunaga, Kyushu Breeding Office for FTBC, Goshi, Japan

Pine wilt disease (PWD) is caused by the pine wood nematode (PWN), *Bursaphelenchus xylophilus*, and remains a very serious problem in pine forests (*Pinus thunbergii* and *Pinus densiflora*) in Japan. A breeding project to develop pine varieties resistant to pine wilt disease was started in 1978 in western Japan, and related projects are promoting throughout Japan. We have been developed 183 varieties in *Pinus thunbergii* and 162 varieties in *Pinus densiflora* that exhibit resistant to pine wilt disease under field conditions until now. One of these varieties in *Pinus thunbergii* was used in self-pollination to develop the population segregating for resistance. Genotyping by sequencing was used to generate single-nucleotide polymorphism (SNP) markers for development of high-density genetic maps and quantitative trait locus (QTL) analyses. One major QTL associated with PWN resistance, distributed on three linkage groups, were discovered in our population. This one major QTL didn't match previously reported PWN resistance QTL. This newly reported QTL found on linkage group 3, which explain between 20 and 30% of the phenotypic variance for PWN, are of particular interest to our breeding program. This QTL provide markers for candidate gene discovery and for future breeding efforts to enhance and pyramid disease resistance.

P0624: Forest Trees

Using Laser Micro-Dissection and qRT-PCR to Analyze Cell Type-Specific Gene Expression in Norway Spruce Phloem

Nina Elisabeth Nagy, Norwegian Institute of Bioeconomy Research (NIBIO), Ås, Norway

The tangentially oriented polyphenolic parenchyma (PP) and radially organized ray parenchyma in the phloem are central in the defense of conifer stems against insects and pathogens. Laser micro-dissection enables examination of cell-specific defense responses. To examine induced defense responses in Norway spruce stems inoculated with the necrotrophic blue-stain fungus *Ceratocystis polonica*, RNA extracted from laser micro-dissected phloem parenchyma and vascular cambium was analyzed using real-time RT-PCR (qRT-PCR) to profile transcript levels of selected resistance marker genes. The monitored transcripts included three pathogenesis-related proteins (class IV chitinase (*CHI4*), defensin (*SPI1*), peroxidase (*PX3*), two terpene synthesis related proteins (*DXPS* and *LAS)*, one ethylene biosynthesis related protein *(ACS)*, and a phenylalanine ammonia-lyase (*PAL*). Three days following inoculation, four genes (*CHI4*, *PAL*, *PX3*, *SPI1*) were differentially induced in individual cell and tissue types, both close to the inoculation site (5 mm above) and, to a lesser degree, further away (10 mm above). These resistance marker genes were all highly induced in ray parenchyma, supporting the important role of the rays in spruce defense propagation. *CHI4* and *PAL* were also induced in PP cells and in conducting secondary phloem tissues. Our data suggests that different cell types in the secondary phloem of Norway spruce have overlapping but not fully redundant roles in active host defense. Furthermore, the study demonstrates the usefulness of laser micro-dissection coupled with qRT-PCR to characterize gene expression in different cell types of conifer bark.

P0625: Forest Trees

Genomic Selection for Growth Traits within and across Four Eucalyptus Breeding Populations

Barbara S. F. Muller[1], Janeo de Almeida Filho[2], Bruno M. Lima[3], Carla C. Garcia[4], Alexandre Missiaggia[5], Aurélio M Aguiar[6], Elizabete K. Takahashi[7], Matias Kirst[8], Orzenil B. da Silva Junior[9], Leandro G. Neves[10] and **Dario Grattapaglia**[9], (1)University of Brasília - EMBRAPA, Brasília, Brazil, (2)State University of North Fluminense, Campos dos Goytacazes, Brazil, (3)FIBRIA S.A. Technology Center, Três Lagoas, Brazil, (4)International Paper do Brasil, Mogi Guaçu, Brazil, (5)Fibria S.A., Jacarei, Sao Paulo, Brazil, (6)Fibria S.A., Aracruz, ES, Brazil, (7)CENIBRA Celulose Nipo Brasileira S.A., Belo Oriente, Brazil, (8)University of Florida, Gainesville, FL, (9)Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil, (10)RAPiD Genomics LLC, Gainesville, FL

Genomic Selection (GS) is a promising approach to accelerate breeding, increase selection intensity and finally integrate genomics into breeding of forest trees. In this study, four unrelated *E. grandis* x *E. urophylla* breeding populations (=758-979, N_e=6.2-19.3) were genotyped and phenotyped for total height (HT) and diameter at breast height (DBH). Prediction models were built with 28,795 to 34,859 high-quality SNPs. The average genome-wide LD (r^2) dropped below 0.2 between 34.8 to 637.7 Kbp. Genomic heritabilities (0.30-0.53) were always lower than the pedigree-based ones, substantiating the inflated estimates of additive genetic variance from pedigree data. Genomic BLUP predictions, 0.36 for HT to 0.67 for DBH, were 5 to 121% better than those obtained by phenotypic selection. Reduced SNP datasets (~3,000 SNPs), including SNPs on single chromosomes (~2,945 SNPs), provided predictive abilities (r_{gy}) equivalent or only slightly less than using all SNPs in these particular populations, showing that prediction is driven largely by relatedness between training and validation sets. The r_{gy} inferred by an additive model using Bayesian methods (BRR, BayesB, BayesA, BayesCπ, BL) reached similar estimates, and a non-additive model (RKHS, machine learning method) improved predictions only slightly 2-10%. When GS models were fitted among populations, the r_{gy} dropped drastically with 0.14 as the highest estimate. These results based on multiples populations and larger sample sizes confirm the encouraging perspective of GS and corroborate earlier predictions that: (1) relatedness is the main driver of GS; (2) the infinitesimal model adequately fits growth data; (3) optimal GS will require population specific models.

P0626: Forest Trees

Accessible Chromatin Mapping in *Eucalyptus grandis* Vascular Tissue

Katrien Brown, Eshchar Mizrachi, Alexander A. Myburg and **Steven G Hussey**, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Pretoria, South Africa

Chromatin architecture changes dynamically during tissue differentiation, influencing DNA accessibility and transcriptional regulation. Functional non-coding genomic elements tend to occur in nucleosome-depleted regions that are hypersensitive to DNase I (DNase hypersensitive sites; DHSs). DNase-seq has aided in annotating novel *cis*-regulatory DNA elements and other functional non-coding regions in Arabidopsis, rice and maize. We aimed to characterize DHSs in developing secondary xylem and phloem of the model hardwood *Eucalyptus grandis*. Chromatin was extracted from developing secondary xylem and phloem tissue from mature field-grown *E. grandis* trees and digested with DNase I to release 50 – 300 bp DNA fragments, Illumina sequencing reads were mapped to the *E. grandis* genome and DHS peaks representing DNase I cleavage hotspots, identified using MACS2, were compared to published RNA-seq and modified histone ChIP-seq data. In developing secondary xylem, the majority of 26 000 biologically reproducible DHS regions (Irreproducible Discovery Rate < 0.05) were intergenic, but genic regions were significantly enriched for DHSs, especially 5' untranslated regions (UTRs) and exons. Relative to genes, DHS density peaked ~200 bp downstream of predicted transcription start sites, and genes containing DHSs within the transcribed or promoter region were significantly more highly expressed than those lacking DHSs. Significant co-occurrences of the DHSs were observed for the activating histone modification H3K4me3 and the repressive histone mark, H3K27me3. We observed local SNP depletion along DHS summits consistent with purifying selection. We highlight the functional relevance of the accessible chromatin profiles, tissue-specific DHS regions and the use of DHS data in transcriptional network reconstruction.

P0627: Forest Trees

***SACPD-C* Mutations Uncover an Impact of Stearic Acid in Leaf and Nodule Structure and Morphology**

Naoufal Lakhssassi[1], Vincent Colantonio[2], Nicholas Flowers[3], Zhou Zhou[1], Jason Henry[3], Shiming Liu[3] and Khalid Meksem[4], (1)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (2)University of Florida, Gainesville, FL, (3)Southern Illinois University, Carbondale, IL, (4)Southern Illinois University Carbondale, Carbondale, IL

Soybean [*Glycine max* (L.) Merr.] is the most widely consumed legume crop in the world, providing 56% of the world's oilseed production (Soystats 2014). Soybean cultivars contain between 3-4% seed stearic acid. Increasing stearic acid confers a higher melting temperature and oxidative stability necessary for solid fat application. Highly-saturated soybean seed oil would be suitable for this end use. Stearoyl-acyl carrier protein desaturase (*SACPD-C*) has been reported to control the accumulation of seed stearic acid; however, no study has previously reported its involvement in leaf stearic acid content and impact on leaf structure and morphology. A subset of an EMS mutagenized population of soybean c.v. 'Forrest' was screened to identify mutants within *GmSACPD-C* gene. Using a forward genetics approach, a nonsense and four missense *Gmsacpd-c* mutants were identified to contain not only high levels of seed, but also high nodule number, in addition to increased leaf and nodule stearic acid content. The EMS nonsense F605 mutant presented the highest seed stearic acid content even reported. Homology modeling and *in silico* analysis of the GmSACPD-C enzyme reveals that most of these mutations were localized near or at conserved residues essential for di-iron ion coordination. Furthermore, mutations at conserved residues cause the highest stearic acid content and correlate with the presence of cell senescence and a necrotic cavity in the nitrogen fixing nodules. Interestingly, soybean plants with GmSACPD-C mutations in non-conserved residues show an increase in stearic acid content and conserving healthy nodules. Thus, random mutagenesis and mutational analysis allows the development of high seed stearic acid content soybeans with no associated negative agronomic characteristics. Finally, results obtained from the current study uncover the impact of GmSACPD-C mutations in leaf and nodule structure and morphology.

P0628: Forest Trees

Genetics Control of 4-Coumarate Coenzyme A Ligase (4CL) Enzyme Involved in Lignin Biosynthesis of European Black Poplar (*Populus nigra* L.)

Bircan Taskiran, Department of Biological Sciences Middle East Technical University, Ankara, Turkey, N. Tulin Guray, Department of Biological Sciences Middle East Technical University, ANKARA, Turkey and Zeki Kaya, Department of Biological Sciences, Middle East Technical University, Ankara, Turkey
European black poplar *(Populus nigra* L.) is considered as one of the most economically significant forest trees with respect to production of wood, pulp, bioenergy and other wood-based products. The content and quality of wood are greatly affected by plant cell wall compositions (cellulose, hemicellulose, and lignin). Although lignin serves as a mechanical barrier as well as protection against pests and pathogens, it emerges as an undesirable polymer for both pulp and bioenergy manufacturing industries because of its by-products which are generated during the removal step of lignin. The 4-Coumarate Coenzyme A Ligase (4CL) is a key regulatory enzyme of the phenylpropanoid pathway that regulates the activation of cinnamic acid, leading to lignin and flavonoid synthesis. With this study, 285 clones with two replications (two ramets per clone) grown in a forest nursery were screened with respect to lignin content and 4CL activity to determine the genetic control of these two traits. The clones highly varied in lignin content and 4CL activity, ranging from 13.24 µg / ml to 48.86 µg / ml in acid soluble lignin, 0.09 to 7.05 units / mg in 4CL specific activity. There is highly significant positive correlation between 4CL activity and lignin content (r = 0.63), whereas low significant negative correlation was found between 4CL activity and diameter growth (r = - 0.22). The results of the study will be further evaluated with respect to the component of total variation attributed to clones, superior lignocellulosic clones and breeding strategies and presented in the poster presentation.

P0629: Forest Trees

Elucidating Transcriptional Regulation Using eQTN Mapping in Populus

Jin Zhang[1], Meng Xie[1], Yongil Yang[1], Kaijie Zheng[1], Lee Gunter[1], Sara S. Jawdy[1], Dave Weston[1], Nancy Engle[1], Timothy J. Tschaplinski[1], Jeremy Schmutz[2,3], Erika A. Lindquist[3], Vasanth Singan[3], Jerry Tuskan[1], Jay Chen[1] and Wellington Muchero[1], (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)DOE Joint Genome Institute, Walnut Creek, CA
Genetic control of transcription was highly complex in higher plants, especially in woody species. Here we are using whole-genome resequencing analysis and transcript expression levels as phenotypes (e-phenotypes) performed a genome-wide association study (GWAS) to reveal the genetic regulate overview in woody model plant *Populus trichocarpa*. The transcriptome data including leaves from 390 *P. trichocarpa* genotypes and xylems from 444 genotypes were used in our eQTN analysis. We have successfully identified novel transcriptional regulators modulating carbon flow between (1) the shikimate and glycolysis pathways and (2) the phenylpropanoid and tryptophan pathways. These serve as new target to redirect carbon flow to competing pathways in order to reduce lignin biosynthesis. Given this success in a short timeframe, we designed the eQTN study to exhaustively identify and characterize transcriptional regulators underlying biomass quality, biofuels and bioproducts yield.

P0630: Forest Trees

Association Analysis of Growth, Disease Resistance, Developmental Properties, and Microbiome Communities, in *Populus deltoides*

Nathaniel Ellis[1], Annette M. Fahrenkrog[1], Christopher Dervinis[1], Leandro G. Neves[2], Marcio F. R. Resende[1], Kelly Balmant[1], Jerald D. Noble[1], Brad Barbazuk[1] and Matias Kirst[1], (1)University of Florida, Gainesville, FL, (2)RAPiD Genomics, Gainesville, FL
Populus deltoides is a short rotation woody crop with strong potential for use in bioenergy production because of its fast growth and wide adaptation to the Midwest and Southern US. In order to uncover genes regulating traits critical for development of a commercial crop we are characterizing a genetically unstructured population of 500 *P. deltoides* individuals. This population was collected from wild stands across its native range, clonally propagated and established in replicated field and greenhouse trials. Individuals were genotyped by sequence capture of all genes expressed in vegetative tissues and over 400,000 polymorphisms have been recorded. Growth and biomass quality measurements have been previously collected from the population growing under greenhouse conditions. This preliminary data lead to a genome wide association study (GWAS) for wood quality traits that uncovered a suite of associated polymorphisms, including a major locus regulating the fraction of syringyl to guaiacyl lignin monomers. The phenotypic and GWAS analysis has expanded to growth properties measured under field conditions (height and diameter), as well as developmental traits critical for selection of elite poplar cultivars. These traits include number of branches, rooting ability, and wood quality and composition. In addition, disease resistance to *Sphaerulina* leaf spot has been quantified, showing moderate to high heritability levels. Resistance assays to *Melampsora* rust (another important pathogen) are ongoing. Finally, we also examined the diversity of the leaf microbiome across the poplar population, for future analysis of its contribution to plant phenotypes. The outcome from these association analyses will be presented.

P0631: Forest Trees

RNAi Suppression of AGAMOUS-like Genes causes Field Sterility in *Populus*

Haiwei Lu[1], Amy Leigh Klocko[1], Amy Brunner[2], Cathleen Ma[1], Anna Carlina Magnuson[1] and **Steven H. Strauss**[1], (1)Oregon State University, Corvallis, OR, (2)Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA
Concerns over transgene dispersal have limited field studies and commercial use of genetically engineered (GE) trees. We seek to mitigate concerns by producing sexual sterility in poplar. Based on gene sequences from *Populus trichocarpa*, we created two RNA interference (RNAi) constructs, PTG and its matrix-attachment-region flanked version MPG, to suppress expression of the two duplicate *AGAMOUS* (*AG*) orthologs in *P. alba* genotype 6K10, an early flowering female clone. A total of 35 transformed events with four ramets per event and 24 wild-type (WT) control trees were planted as part of a larger field trial in 2011. Six out of 22 flowering PTG events and 11 out of 12 flowering MPG events showed a modified floral phenotype; their floral buds flushed early in the field and the capsules on each catkin often had “carpel-inside-carpel” phenotypes as expected from impairment of AG activity. A complete disruption of ovule and seed production was observed in a number of gene insertion events within both constructs. We also discovered suppression at two AG-like genes (*AGL11*), in sterile events. In all

cases, trees appeared normal in their vegetative morphology and growth, and alterations in floral phenotypes were stable over multiple years. RNAi suppression of *AG*-like genes appears to be a safe and effective means of genetic containment in poplar.
We thank the USDA Biotechnology Risk Assessment Grants (no. 2010-33522-21736 and no. 2011-68005-30407) and the Tree Biosafety and Genomics Research Cooperative at OSU for support.

P0632: Forest Trees

Pan-Genome of Silver Birch (*Betula pendula*)

Sitaram Rajaraman[1], Victor A. Albert[2] and Jarkko T Salojärvi[1], (1)University of Helsinki, Helsinki, Finland, (2)University at Buffalo, Buffalo, NY

Silver birch (*Betula pendula*) is an important pioneer tree species that grows in boreal forests across Europe and Asia. We recently assembled the reference genome and carried out a population genomics study on the species. Selective sweeps were found acting mostly on the genomic regions originating from whole genome duplications, whereas most recent tandemly duplicated genome regions were depleted of sweeps. Here we study the standing pool of genomic structural variation among silver birches by estimating the pan-genome of silver birch from individuals sampled from 12 different locations across Eurasia. The individual-specific reads were assembled into scaffolds using SPAdes assembler. After removing organellar and bacterial contaminants the *de novo* assembly added 6 Mbp of contigs per individual, containing approximately 2000 gene fragments predicted by AUGUSTUS.
Tandemly duplicated genes were enriched among the individual-specific genomes, corroborating recent results that copy-number variation forms the largest pool of standing variation among individuals. Gene ontology categories related to environmental responses were found to be enriched among the dispensable part of the genome.
References: Salojärvi, Smolander, Nieminen, Rajaraman et al. Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nature Genetics* **49**:904–912 (2017).

P0633: Forest Trees

Dissecting Genetic Resistance to Willow Leaf Rust (*Melampsora spp.*) using Common Parent Mapping Populations of Shrub Willow (*Salix spp.*), a Biomass Energy Crop

Dustin G. Wilkerson[1], Craig H. Carlson[2], Chase R. Crowell[2], Christine D. Smart[2] and Lawrence B. Smart[2], (1)Cornell University, Ithaca, NY, (2)Cornell University, Geneva, NY

Willow leaf rust (*Melampsora spp.*) is a devastating plant pathogen that can be responsible for up to 100 percent defoliation in cultivated shrub willow (*Salix spp.*) grown as a biomass energy crop, significantly impacting yield. In order to map resistance to willow leaf rust and to serve as a genetic resource for mapping other traits of interest including yield components, insect resistance, and physiological traits, we developed species hybrid F_1 mapping populations with a common parent of *Salix purpurea*. We produced eight families with either *S. purpurea* 94006 as the female parent (the reference genome) or *S. purpurea* 94001 as the male parent crossed to individuals of *S. suchowensis, S. viminalis, S. udensis, S. integra, S. koriyanagi,* or *S. alberti*. These mapping populations were planted in adjacent field trials in Geneva, NY with 88 to 150 individuals per family in each of four randomized complete blocks per trial. These progeny were genotyped using genotyping-by-sequencing and genetic maps were developed from segregating SNP markers. During the summer of 2017, ratings were collected for damage from imported willow leaf beetle (*Plagiodera versicolora*), potato leafhopper (*Empoasca fabae*), and severity of willow leaf rust, as well as leaf SPAD and specific leaf area measurements. QTL mapping for each of these traits will be presented.

P0634: Forest Trees

Haplotype Frequency Analysis of the Turkish *Salix* L. (Salicaceae)

Pelin Acar[1], Funda Ö Degirmenci[1], Zeki Kaya[1] and **Bircan Taskiran**[2], (1)Department of Biological Sciences, Middle East Technical University, Ankara, Turkey, (2)Department of Biological Sciences Middle East Technical University, Ankara, Turkey

The genus *Salix* L. is represented by more than 500 species in the world where 27 species are naturally found in Turkey. This genus has been commonly used as biomass production for energy, phytoremediation and pharmacological products. The objective of this study is to estimate the haplotype frequency analysis of native willow species in Turkey. The sequence data are obtained from non-coding (*trn*L) and coding (*mat*K and *rbc*L) cpDNA gene regions of *Salix* species (24 species and one hybrid). Results showed that all haplotype trees gave the similar outcomes separating haplotypes into two clusters (subgenera; *Salix* and *Vetrix*). The most diverse haplotype in *mat*K region (23 haplotype) are identified in Turkish willows. Haplotype sharing was increased for both *trn*L and *rbc*L gene regions, but they were still high (13). Incomplete lineage sorting and introgressive hybridization are the phenomena that likely determine the haplotype composition of Turkish *Salix* sp. Haplotype sharing is common within the geographically close members of subgenus (*Salix* or *Vetrix*) but not between two subgenera of *Salix*. Anatolia's varied regional and specific climate seem to create suitable environments for haplotype sharing within subg. *Vetrix* species (higher altitude and cooler climates) and within subg. *Salix* species (warmer climate). Furthermore, analysis of molecular variance (AMOVA) and pairwise F*st* values for Turkish *Salix* species, Old World and New World *Salix* based on cpDNA *mat*K and *trn*L gene regions revealed that the high variation within groups and differences between Turkish subgenera are consequences of contribution of OWS and NWS.

P0635: Fruit Species

NRSP10 Update: Development of Crop Genomics, Genetics and Breeding Database Resources

Dorrie Main[1], Sook Jung[1], Chun-Huai Cheng[2], Taein Lee[3], Jodi L. Humann[1], Jing Yu[1], Heidi Hough[1], Morgan Frank[1], Kristin Scott[1], Katheryn Buble[4], Deah McGaughey[1], Victor Unda[1], Yu Ma[1], Jim McFerson[5] and Cameron Peace[1], (1)Washington State University, Pullman, WA, (2)Washington State University, Pullman, Pullman, WA, (3)WSU, Pullman, WA, (4)Washington State University, Pulmman, WA, (5)Washington State University, Wenatchee, WA

National Database Resources for Crop Genomics, Genetics and Breeding Research is a US Land Grant Universities, USDA NIFA and industry funded project which provides standardized database and informatics resources for undeserved or specialty crops such as tree fruit, nuts, and berries. It builds on existing database resources developed for Rosaceae (Genome Database for Rosaceae, www.rosaceae.org), Citrus (Citrus

Genome Database, www.citrusgenomedb.org), Vaccinium (Genome Database for Vaccinium, www.vaccinium.org), Cool Season Food Legumes (Cool Season Food Legume Genome Database, www.csfl.org) and Cotton (CottonGen, www.cottongen.org). Developed using Tripal, an open-source, resource-efficient, modular, well supported platform, these community databases provide centralized access to integrated genomic, genetic and breeding data and analysis tools for 24 crops representing a combined annual production value of over $25 B. We highlight the latest data and functionality provided in these databases and plans for future development and sustainability.

P0636: Fruit Species

Quickly and Painlessly Obtaining Phased Genome-Wide SNP Data of Fruit Breeding Individuals using previously Curated Haplotype Information

Stijn Vanderzande[1], Lichun Cai[2], Cassia Da Silva Linge[3], Nicholas Howard[4], Amy Iezzoni[2], Ksenija Gasic[3], Eric van de Weg[5] and Cameron Peace[1], (1)Washington State University, Pullman, WA, (2)Michigan State University, East Lansing, MI, (3)Clemson University, Clemson, SC, (4)University of Minnesota, St Paul, MN, (5)Plant Breeding Wageningen University & Research centre, Wageningen, Netherlands

Taking full advantage of genome-wide DNA information by breeding programs and allied scientists requires quick obtainment of high-quality genotypic data. SNPs are abundant, but acquiring accurate SNP data is not simple. Previously, phased SNP-based haplotypes for hundreds of cultivars and breeding individuals were created in RosBREED using a pipeline that efficiently identified errors while retaining as many informative loci as possible. These datasets encompassed 3863 and 1617 SNPs combined in 917 and 196 haploblocks for 830 apple and 528 sweet cherry individuals, respectively, while haploblock analysis for 4005 SNPs in 620 peach individuals is ongoing. Despite detecting genotype-call errors efficiently, the pipeline required considerable time to obtain final datasets. Anyone desiring such genome-wide SNP information on new individuals faces the daunting task of repeating laborious curation efforts. However, new individuals are typically related to RosBREED's baseline germplasm, which enables curation shortcuts. We developed a method to obtain high-quality, genome-wide, phased haplotypic data quickly for newly genotyped germplasm. First, a robust SNP set requiring no curation was identified. Next, possible haplotypes for new individuals were deduced by comparing their robust-SNP genotypes to possible combinations of baseline parental haplotypes. Where haplotype assignment was ambiguous, baseline patterns of linked haplotypes were used. Finally, genotypes for remaining SNPs were imputed from haplotypes. This method was able to determine haplotypes from raw SNP data for new individuals in weeks rather than months. Genome-wide genetic marker information can now be readily obtained for close relatives of previously haplotyped germplasm, facilitating routine integration into breeding decisions.

P0637: Fruit Species

Predict: A Tool for Predicting Progeny Distribution in New Crosses for Multiple Correlated Traits

Julia Piaskowski[1], Craig Hardner[2] and Cameron Peace[1], (1)Washington State University, Pullman, WA, (2)University of Queensland, Brisbane, Australia

Plant breeding is widely characterized as a "numbers game" in which thousands of genetically distinct individuals are generated, largely through hybridization, and most are discarded. This initial step is particularly costly for tree fruits because they are expensive to grow and have a long juvenility period prior to phenotypic evaluation. Selection of superior individuals in a population is most efficient when population means are favorable and the genetic variation is sufficiently large for the target traits. Currently, there are no known tools for predicting the progeny distribution of a cross for clonally-propagated outcrossing crops. Furthermore, existing methodologies developed for other species are for single traits only. We have developed an online tool for predicting the distribution of progeny from a biparental cross across multiple correlated traits. Using breeding values of prospective parents, the genetic variance/covariance of a breeding population, and the expected inbreeding coefficient between two parents, an F1 population is simulated assuming the multivariate normal distribution. Two approaches were evaluated for predicting the genetic variance of a new cross: expected covariance among relatives and *in silico* segregation of alleles. The accuracy of this tool in predicting progeny distributions was tested on sweet cherry (*Prunus avium*) biparental crosses. *PREDICT* is intended to help breeders of highly heterozygous crops find cross combinations which maximize the probability of producing a superior individual, thus improving crossing efficiency.

P0638: Fruit Species

Internal Standards Roulette: Best Practices for HS-SPME-GC-MS Volatile Analyses in Plant Populations

Elizabeth A. Burzynski[1], Bruce Reisch[2], Itay Gonda[3], Majid R. Foolad[4], James J. Giovannoni[3,5] and Gavin Sacks[1], (1)Food Science, Cornell University, Ithaca, NY, (2)School of Integrative Plant Science, Cornell University, Geneva, NY, (3)Boyce Thompson Institute for Plant Science, Ithaca, NY, (4)Plant Science, Pennsylvania State University, University Park, PA, (5)USDA-ARS Robert W. Holley Center, Ithaca, NY

Headspace-solid-phase micro-extraction (HS-SPME) coupled to GC-MS (gas chromatography-mass spectrometry) is widely used for convenient analysis of volatiles in plant populations, such as mapping populations. HS-SPME is well-known to suffer from matrix effects, and a common practice to compensate for such effects is to use a single surrogate standard, with the assumption that differences in relative matrix effects for any pair of compounds is small among individuals. We demonstrate that this assumption is not valid using two different plant mapping populations spiked with a cocktail of isotopically labeled standards. Because these standards are non-native, relative ratios between HS-SPME-GC-MS signals should be consistent across individuals if there were no Individual × Standard dependent matrix effects. However, in an interspecific grape (*Vitis* sp.) mapping population, we show that relative HS-SPME-GC-MS responses of any two internal standard pairs ranged from 17% to 249% RSD (relative standard deviation). In a follow-up study, using samples from a tomato recombinant inbred line (RIL) population, derived from a cross between *Solanum lycopersicum* breeding line NC EBR-1 and *S. pimpinellifolium* accession LA2093, we observed variation in relative responses of 6% to 64% RSD for pairs of standards. We propose a *post-hoc* strategy to characterize the extent of Individual × Standard dependent matrix effects and to identify well-suited surrogate standards.

P0639: Fruit Species

Evaluating the Effect of Ploidy Level and Non-Additive Effects in Autotetraploids Genome-Wide Prediction

Ivone de Bem Oliveira[1], Marcio F. R. Resende[1], Luis Felipe V. Ferrão[1], Juliana Benevenuto[1], Rodrigo Rampazo Amadeu[2], Jeffrey Endelman[3], Catherine Cellon[1,4], James Olmstead[1,5], Matias Kirst[1] and Patricio R. Munoz[1], (1)University of Florida, Gainesville, FL, (2)University of São Paulo, Piracicaba, SP, Brazil, (3)University of Wisconsin, Madison, WI, (4)Duda Farm Fresh Foods,, Oviedo, FL, (5)Driscoll's, Watsonville, CA

Polyploidy events are not an exception in plants, about 70% of the Angiosperms and about 95% of the Pteridophytes have suffered at least one polyploidization event. This phenomenon is considered a powerful evolutionary source, since new phenotypic variations can be generated. Our goal is to investigate the effect of ploidy in the estimation of genetic variance components and on prediction of phenotypes. The ultimate goal is to accelerate autotetraploid breeding by improving the capacity to predict complex traits with genome-wide selection models. For this, a blueberry breeding population of 1847 individual were phenotyped for yield and fruit traits and genotyped using sequence capture (~86Kb SNPs). Besides considering ploidy, Genome Selection models implemented also considered dominance effects. No significant improvements in breeding values predictions were found when comparing diploid with tetraploid models. However, for most traits the maker-based models considering tetrasomic inheritance with the insertion of nonadditive effects generates a better fit than disomic inheritance models, even when an extensive pedigree data was used. The insertion of nonadditive effects improved breeding value prediction. Furthermore the separation of estimated additive and nonadditive genetic variance was better when marker-based genomic relationships were used. This novel result improves our current understanding of the genetic control of quantitative traits in this autotetraploid species. Further studies may also generate gains not only to improve selection, but also generating information about the genome of the crop, and bringing new information about the influence of ploidy in the phenotypic expression.

P0640: Fruit Species

The Split Personality of the Clonally Propagated Grape Genome

Susan Thomson[1], Tim Millar[1], Darrell Lizamore[2], Philippa Barrell[1], **Samantha Baldwin**[1] and Chris Winefield[2], (1)The New Zealand Institute for Plant & Food Research LTD, Christchurch, New Zealand, (2)Lincoln University, Christchurch, New Zealand

Grape (vitis vinifera) is a domesticated crop that has been clonally propagated for centuries. The varietal Pinot noir is one of the oldest, grown for wine production for at least 7-800 years. Chimerism and accumulation of genome mutations is known to exist within clonally propagated grape vine and have contributed towards the identification of distinct cell layers. To date this has been shown through the application of low resolution SSR markers to selected loci. Through the use of callus cell culture derived from diploid anther material, we have regenerated several plants. We present whole genome variant analysis as well as transposon fingerprinting using the tool TEfingerprint, of parental meristem and 6 regenerants highlighting distinct genome regions of high divergence between cell layers.

P0641: Fruit Species

Grape Berry Ripening: Gene Expression to Aroma Analysis

Padmapriya Swaminathan[1], Anne Fennell[1], Kalley Besler[1], Somchai Rice[2], Jacek Koziel[2], Zata Vickers[3] and Emily Del Bel[3], (1)South Dakota State University, Brookings, SD, (2)Iowa State University, Ames, IA, (3)University of Minnesota, St. Paul, MN

Frontenac and Marquette are cold climate cultivars emerging from *Vitis vinifera* and North American Vitis species. Chemical composition of *Vitis vinifera* cultivars have been extensively studied but not well understood for these cold hardy cultivars. Titratable acidity (TA), pH and soluble solids indicate grape maturity and help determine harvest time for most growers and winemakers. Characterizing the grape berry ripening profile of Frontenac and Marquette berry pulp and skin through transcriptomic, sensory and flavor analyses is critical to understand and obtain the optimal balance between sugars, acids and flavor to aid in harvest decisions and hence better wine quality. Transcriptomic analysis shows changes in gene expression impacting Anthocyanins, Terpenoids, Flavonoid biosynthesis pathways.

P0642: Fruit Species

Genetic Architecture of Complex Traits in Table Grape, a First Glance of the Principal Determinants of Fruit Quality: Berry Weight and Firmness.

Nilo Mejía, Nallatt Ocarez, Reynaldo Núñez, Nicolás Jiménez, Jorge Lagrèze, Ignacio Morales and Bruno G. Defilippi, Instituto de Investigaciones Agropecuarias, Centro Regional La Platina, Santiago, Chile

For table grape breeding, assisted selection for complex traits was underestimated due to scarce, not reproducible nor reliable quantitative trait loci (QTL) results based generally in small progenies and poor genome coverage. To overcome such limitations we developed a large F1 bi-parental progeny (Muscat of Alexandria x Crimson Seedless, n > 600) that was characterized at genetic level with 5,279 SNPs and for berry weight and firmness over three seasons. In order to identify genotype-phenotype associations within this progeny we performed three complementary methodologies: fine QTL mapping, genome wide association study (GWAS) and genomic selection (GS) based in a ridge regression best linear unbiased prediction model (rrBLUP). QTL mapping results reveal the existence of more than 15 minor responsible *loci* for each trait, reflecting the very complex genetic nature of berry weight and firmness. The combined use of GWAS and QTL analysis points at 4 to 6 significant and reproducible QTLs for each of the targeted traits. For each trait, a selection method that considers the combined use of these 4 to 6 most significant QTLs was proposed and called “haplotype-based selection” which explains up to 60 and 34 % of the phenotypic variation for berry weight and firmness respectively. For the same traits, accuracy of GS determined with the validation population reaches 71 and 48 % respectively. Together, these results show that assisted selection of very complex traits could be performed with a small panel of markers rather than yet expensive whole-genome genotyping dependent GS.

Financed by FONDEF G09i1007, FONDECYT 1170586 and Biofrutales S.A.

P0643: Fruit Species

Identification of Haplotypes Controlling Seedless By Genome Resequening of Grape

Soon-Chun Jeong, Korea Research Institute of Bioscience and Biotechnology, Chungbuk, Korea, Republic of (South) and Youn Young Hur, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wanju, Korea, Republic of (South)
Seedless is an important trait that determines eating quality of fresh grape fruits. In order to establish a facile marker-assisted selection system for this seedless trait by identifying haplotypes of seedless-controlling loci, we have resequenced more than 20 grape accessions with higher than 50x genome coverage. We is currently analyzing the sequencing data and will present general polymorphism patterns of grape as well as haplotype networks of seedless-controlling loci.

P0644: Fruit Species

Image-Based Phenotyping and Genetic Control of Cluster Density Traits in Interspecies Grapevine Hybrids

Anna Nicole Underhill, Cory D. Hirsch and Matthew D. Clark, University of Minnesota, St. Paul, MN
Cluster compactness is an economically important trait in grapevine (Vitis spp.), as increased fruit compactness has been correlated to increased fungal, bacterial, and insect infections that result in fruit loss. Quantifying cluster compactness using current phenotyping methods is difficult, as they rely on trained visual assessment or time-consuming physical measurements. Here, we captured fruit cluster images from 123 F_1 individuals and segmented them based on color, then extracted values for cluster length, width, and compactness, plus berry weight, size, and color. Additionally, physical bunch measurements were taken to compute density based on an index of cluster weight divided by the square of cluster length. Mean cluster weight in the population was 42.11g with a standard deviation of 19.49g, and mean cluster length was 10.94cm with a standard deviation of 2.60cm; the parental genotypes had mean weights of 54.14g and 89.68g and mean lengths of 8.19cm and 15.5cm, respectively. Cluster compactness had a mean of 0.36 g/cm^2 and a standard deviation of 0.13 g/cm^2 with the parents averaging 0.52 g/cm^2 and 0.35 g/cm^2. Cluster weight and cluster length were found to be positively correlated, with a correlation coefficient of r = 0.99, while both were negatively correlated to cluster compactness, r = -0.94 for weight and r = -0.97 for length. To better understand genetic control of fruit components, QTL mapping for each trait was carried out and a significant QTL for 20 berry weight accounting for approximately 14.68% of variation was found on linkage group 9.

P0645: Fruit Species

ABA Sensitivity in Four *Vitis* Species

Haley S. Toups, University of Nevada, Reno, Reno, NV and Noe D. Cochetel, Dept of Biochemistry & Molecular Biology, University of Nevada, Reno, Reno, NV
Vitis vinifera is a significant agricultural crop; *Vitis vinifera* cultivation and processing provides numerous products including wine and juice, tens of thousands of jobs, and a $34 billion industry in California alone. *Vitis vinifera* cultivation is dependent on grafting desirable fruit scions onto phylloxera resistant root stocks, which have been demonstrated to play a significant role in fruit production and abiotic stress tolerance. Understanding traits of rootstocks is crucial to optimizing fruit production in adverse conditions.
Abscisic acid (ABA) is a key plant hormone involved in the regulation of many physiological processes occurring in response to abiotic stresses including water deficit, salinity, and cold tolerance. ABA mediates seed germination, dormancy, and stomatal closure, which directly affects photosynthesis and fruit production. Cultivar specific stomatal responses to water deficit in *Vitis vinifera* has led to a spectrum of classifications ranging from isohydric to anisohydric.
Transcriptomic and physiological responses to drought and ABA will be reported. Stomatal responses vary between Vitis species*: V. Riparia, V. vinifera, V. champinii*, and *V. vinifera x girdiana* hybrid, exposed to identical water deficit and control conditions. This species-specific response is hypothesized to be due to differences in ABA concentrations and sensitivities. This research demonstrates the different stomatal responses between species may be due to an intrinsic difference between [ABA], ABA metabolism, and ABA signaling needed to illicit drought responses in Vitis. Understanding ABA sensitivity in Vitis rootstocks will allow for optimized fruit production in stressful conditions by planting appropriate scion/rootstock graft combinations based on environmental conditions.

P0646: Fruit Species

Using Genetic Variability of Grapevine to Study Drought Tolerance through Genotype-Specific Transcriptome Signatures during a Water Deficit

Noe D. Cochetel[1], Richard L. Tillett[2], Karen A. Schlauch[2] and Grant R. Cramer[1], (1)Dept of Biochemistry & Molecular Biology, University of Nevada, Reno, Reno, NV, (2)Nevada Center for Bioinformatics, University of Nevada, Reno, Reno, NV
Drought is a major abiotic stress that drastically limits crop production and global warming will exacerbate it. This highlights the urgent need to adapt our current agricultural practices and modify our crop cultivars. These statements are particularly relevant in the semi-arid Southwestern region of the United States. In that context, Nevada is a near perfect place for conducting research on water deficit and can help us to reveal the upcoming impacts of the global warming. The model species of this project, grapevine, is a very important fruit crop in the USA and all around the world, and is known to present genetic variability concerning its stress tolerance in arid and semi-arid environments. The present work aimed to investigate the molecular basis of drought tolerance mechanisms in grapevine including species native to Nevada. An RNA-seq analysis was performed on four different genotypes (*Vitis vinifera*, *Vitis riparia*, *Vitis champinii* and *Vitis vinifera* x *girdiana* hybrid) in response to a drought stress applied during two weeks. Results showed that a core response to water deficit is shared between genotypes. Interestingly, a Weighted Gene Correlation Network Analysis performed on these data revealed distinct genotype-specific modules (gene clusters). These results are even more interesting since each of the selected genotypes are known for their different levels of drought tolerance. Thus, the analysis of these modules could lead to the identification of candidate genes for marker-assisted breeding of grapevine rootstock in order to improve their drought tolerance and will reinforce our current knowledge on drought.

P0647: Fruit Species

Identification of Nucleotide Binding Site Leucine Rich Repeats (*NBS-LRR*) Genes Associated with Fungal Resistance in *Vitis vinifera*

Neetu Goyal[1], Garima Bhatia[1] and Kashmir Singh[2], (1)PANJAB UNIVERSITY, CHANDIGARH, India, (2)Panjab University, Chandigarh, India
Nucleotide binding site leucine rich repeats (NBS-LRR) encompass a wide class of disease resistance (R) proteins that play an extensive role to protect plants during pathogen attack. In grapevine (*Vitis vinifera*), viticulture productivity is on immense loss due to powdery mildew (PM) infection and research is going on to find resistance loci or resistance/tolerance imparting genes against PM. Presently, we performed genome-wide study and identified 386 *NBS-LRR* genes in grapevine. Next, sequence characterization of respective genes was done on the basis of conserved protein domains, chromosomal locations, and functional annotation by Gene Ontology (GO) mapping. A major occurrence of coiled coil (CC) as compared to Toll/interleukin-1 receptor (TIR) has been depicted in Domain analysis of NBS-LRR proteins. Chromosomal mapping of *NBS-LRR* genes indicated a major clustering over chromosomes 9, 12, 13, and 19. In Blast2GO study, majority of genes were found to have function in defense against biotic stresses. We also identified PM-responsive *NBS-LRR* genes by performing normalized expression analysis (FPKM) of 386 *NBS-LRR* genes in three different varieties of grapevine; namely Carignan, Karadzhandal, and O34-16. Further, validation of these genes was done by quantitative real-time PCR (qRT-PCR). Consequently, we identified 30, 28, and 30 differentially expressed PM-responsive *NBS-LRR* genes in Carignan, Karadzhandal, and O34-16 varieties, respectively. The evolutionary relationships amongst the *NBS-LRR* genes were obtained based on Phylogenetic analysis. Altogether, we identified PM-responsive genes that would act as potential resource to improve the disease resistance in grapes.
KEYWORDS: Grapevine, fungal infections, powdery mildew resistance, *NBS-LRR* genes, expression

P0648: Fruit Species
***Vitis sylvestris* Harbours Genetic Variation in Stilbene Metabolism**
Dong Duan, Northwest University, Xi'an, China
Vitis vinifera L. ssp. *sylvestris* (Gmelin) Hegi, the European Wild Grape and ancestor of cultivated grapevine varieties (*V. vinifera* L. ssp. *vinifera*) is the only wild grapevine species existing in Europe. This important Crop Wild Relative (CWR) species is almost extinct and persists only in residual habitats. Recently, CWR have attracted the attention as valuable genetic resources for breeding. Some of the *sylvestris* genotypes harbour valuable resistence factors against several diseases of grapevine, such as Powdery Mildew (*Erysiphe necator*), Downy Mildew (*Plasmopara viticola*), and Black Rot (*Guignardia bidwellii*). However, since they had not been previously exposed to these diseases before, the resistance must be different from conventional effector-triggered immunity. Stilbenes are a small family of plant secondary metabolites, which have implications for plant basal immunity and human health. And they are generally involved in the response to biotic and abiotic stresses. In this study, we showed that some of the resistant *sylvestris* genotypes had elevated levels of stilbenes in response to different forms of stress, correlated with elevated induction of metabolic genes in the stilbene synthesis pathway.
Keywords: Defence, genetic diversity, grapevine (*V. sylvestris*), stilbenes

P0649: Fruit Species
High-Density Linkage Maps and Loci for Fruit Color and Flower Sex in Muscadine Grape (*Vitis rotundifolia*)
Jennifer Lewter[1], **Margaret L. Worthington**[1], John Clark[1], Chris Owens[2], Patrick Conner[3] and Aruna Varanasi[1], (1)University of Arkansas, Fayetteville, AR, (2)IFG, Bakersfield, CA, (3)University of Georgia
Muscadine grapes, *Vitis rotundifolia* Michx. subgenus *Muscadinia* Planch. (2n=40), are a disease-resistant specialty crop native to the southern United States. Although muscadines are related to 'bunch' grapes in the subgenus *Euvitis* Planch. (2n =38), they have smaller clusters, thicker skins, unique fruity/floral aromas, and relatively unstable diglucoside anthocyanin profiles. High-density linkage maps of two F_1 muscadine populations ('Supreme' x 'Nesbitt' and 'Black Beauty x 'Nesbitt') were developed using genotyping-by-sequencing (GBS). Both populations segregated for flower type (perfect vs. female flowers) and berry color (black vs. bronze). The linkage maps were each composed of 20 linkage groups, with 2019 and 1209 total markers respectively. Synteny was strongly conserved between the muscadine linkage maps and the *V. vinifera* physical map. The division of *V. vinifera* chromosome 7 into two independently segregating linkage groups accounted for the higher chromosome number in muscadine. The locus controlling flower type in muscadine mapped to a region spanning 4.6-5.4 Mbp on chromosome 2, which includes the previously described *V. vinifera* subs *sylvestris* sex locus. While the VvMYBA1 transcription factor controlling fruit color in *V. vinifera* is located on chromosome 2, the muscadine fruit color locus mapped to a region spanning 11.1-11.9 Mbp on chromosome 4. This finding suggests that a mutation in a different gene in the anthocyanin biosynthesis pathway differentiates between black and bronze berry color in muscadine. These dense linkage maps lay the groundwork for marker-assisted breeding in muscadine and provide insight on the evolution of *Vitis* species.

P0650: Fruit Species
Rootstock and Scion Interaction: miRNA Perspective?
Muhammad Umair Ahsan[1], Alice Hayward[1], Christine Beveridge[2] and Neena Mitter[1], (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia, (2)School of Biological Sciences, The University of Queensland, St. Lucia, Australia
Grafting is widely used in most horticultural tree crops, such as avocado (*Persea Americana*), for commercial orchard production. It is a technique in which a shoot (scion) is combined to the roots (rootstock) of another plant and then grown together. Avocado rootstocks can be either seedling plants (juvenile origin) or rooted tree cuttings (mature origin). Based on phenotypic observation it is generally considered that the scion is the main contributor to grafted plant maturity, however, it is not known if the rootstock has any influence on shoot maturation in avocado scions. Recent studies indicate graft transmissibility of microRNAs (miR156 and miR172), which are key regulators of plant development in other plant species. In this study, the activity of miR156 and miR172, and their putative targets, was quantified in avocado scions grafted to both seedling and clonal rootstocks. Pre-grafting and post-grafting samples were collected at various times from 3-months to 2 years post-grafting. We show that miR156 abundance in the scion is scion-dependent and not affected by the type of rootstock. On the other hand, the rootstock may have an impact on miR172 abundance in the scion, consistent with findings in other plants suggesting its graft transmissibility. Target gene quantification *(SPL4, SPL9 and AP2)* revealed that *PaSPL4* and *PaAP2a* were expressed antagonistically with

respect to miR156 and miR172, respectively. These results suggest that the scion is largely responsible for avocado grafted tree miR156 levels, however, the juvenility of the rootstock (seedlings vs clonal) may contribute to inter-graft regulation on miR172.

P0651: Fruit Species

Strawberry (*Fragaria*) Genomic Resource Development Update: UNH

Thomas M. Davis, Yilong Yang, Melanie Shields, Lise L. Mahoney and Dan Frailey, University of New Hampshire, Durham, NH

We are developing genomic and germplasm resources, knowledge, and genotyping tools in support of research and marker-assisted breeding in strawberry (Fragaria). Status reports will be presented on the following projects: the new diploid (Fragaria iinumae) reference genome; SNP mapping in diploid F. vesca; a novel approach to genome assembly in an octoploid; marker-assisted cultivar development.

P0652: Fruit Species

Genotyping-by-Sequencing and Reference Genome Enabled Variant Discovery in Octoploid Strawberry

Mitchell J. Feldmann[1], Michael A Hardigan[1], Thomas Poorten[1], Charlotte B. Acharya[1], Marivi Colle[2], Robert Van Buren[2], Patrick P. Edger[2] and Steven J. Knapp[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Department of Horticulture, Michigan State University, East Lansing, MI

Genotyping-by-sequencing (GBS) approaches have enabled routine high-density genome-wide DNA variant discovery in numerous agriculturally important species. Applications of GBS in octoploid (2n = 8x = 56) strawberry (*Fragaria* × *ananassa*) have been hindered by the absence of a reference genome for physically mapping DNA sequences; for discovering variants with sub-genome resolution, or effectively distinguishing homologous from homeologous variation. High-quality reference genome assemblies have recently emerged, supplying the foundation for this study, which focused on demonstrating the utility of GBS for calling sub-genome specific DNA variants in octoploid strawberry. To reduce genomic DNA complexity, double-digest protocols were tested on diverse accessions with two restriction enzyme combinations (*PstI-MspI* and *HindIII-MspI*). GBS libraries were sequenced on an Illumina HiSeq 4000 using a 150 bp paired-end protocol. For the purpose of this study, we describe the deployment of a flexible bioinformatic pipeline for GBS-facilitated variant discovery in octoploid strawberry. The percentage of uniquely mapped reads ranged from 51.41% for *PstI-MspI* to 55.56% for *HindIII-MspI* resulting in 1,591,764 and 2,362,556 unique locations, respectively. The number of discovered variants was 2.5-fold greater for *HindIII-MspI* (491,811) than *PstI-MspI* (199,486). The GBS protocols uncovered a dense genome-wide landscape of DNA variants for high-precision genetic mapping, identification of DNA variants associated with agriculturally important phenotypes, genomic-enabled breeding, and other applications in octoploid strawberry.

P0653: Fruit Species

Domestication History of Strawberry: Population Bottlenecks and Restructuring of Genetic Diversity through Time

Michael A Hardigan, Department of Plant Sciences, University of California, Davis, Davis, CA

Domesticated strawberry (*Fragaria* x *ananassa*) is an octoploid hybrid (*2n=8x=56*) of Chilean landrace beach strawberry (*F. chiloensis* [L.] Miller) and North American common strawberry (*F. virginiana* Miller). University of California strawberry varieties represent an elite genetic architecture within the *ananassa* hybrid complex, accounting for 60% of worldwide consumption, and are recognized for their value in developing cultivars of high yield and excellent fruit quality. Allopolyploidy historically restricted exploration of genetic diversity, selection and bottlenecks that produced modern cultivars. A panel of 1,300 octoploid *Fragaria* genotypes containing 100 wild accessions from North and South America, 157 heirloom cultivars, and the University of California (UC) germplasm collection was fingerprinted on the Axiom iStraw35 array to investigate population restructuring following strawberry domestication and identify targets of selection using 16,492 subgenome-specific markers anchored to an improved diploid assembly. This study reports previously unknown features of the octoploid genome, including preferential retention of the *F. virginiana* wild progenitor background. It is revealed UC strawberry genetics underwent a bottleneck creating a modern *F. x ananassa* population genetically distinct from cultivars bred elsewhere in North America and Europe, resulting in greater divergence between modern UC and non-UC *F. x ananassa* (FST=0.207) than non-UC cultivars and strawberry's wild progenitors (FST=0.182). Selection of photoperiod-insensitive flowering is shown to be a driver of this bottleneck, including transitioning from complex polygenic sources of day-neutrality to a simpler mechanism involving dominant QTLs on a single chromosome arm, independent of the *TERMINAL FLOWERING1*-mediated source identified in diploid *Fragaria* species.

P0654: Fruit Species

A Genetic Linkage Map for *Fraxinus pennsylvanica* and Synteny Analysis with Asterid and Rosid Species

Di Wu[1], Jennifer Koch[2], Mark Coggeshall[3], Nicole Zembower[1] and John E. Carlson[1], (1)Dept. of Ecosystem Science & Management, Pennsylvania State University, University Park, PA, (2)The U.S Forest Service, Delaware, OH, (3)University of Missouri, Columbia, MO

Green ash (*Fraxinus pennsylvanica*) is an outcrossing, diploid (2n=46) hardwood tree species. The rapid invasion of emerald ash borer (EAB) from Asia is threatening all native ash species in North America. Green ash, the most widely distributed ash species, is being severely affected by EAB infestation, yet few genome-wide resources for genetic studies and improvement of green ash are available. In this study, we constructed the first genetic linkage map for green ash. We identified thousands of single nucleotide polymorphisms (SNPs) across the genome through genotyping-by-sequencing (GBS) and screened hundreds of genomic- and EST-based microsatellite markers (SSRs) from previous *de novo* assemblies (Staton et al, 2015; Lane et al. 2016). Segregation analysis was conducted using 2,719 high quality SNP and 84 SSR polymorphic markers in a full-sib family of 90 individuals. The female parent genetic map consisted of 992 markers at 760 distinct positions spanning 1,562.64cM, with an average marker interval of 2.22cM. The male parent map consisted of 755 markers on 1,744.12cM, with an average marker interval 2.87cM. A consensus genetic map was produced with a total of 1,201 segregating SNP and SSR markers on 23 linkage groups spanning 2008.87cM at an average inter marker distance of 1.67 cM, with a minimum logarithm of odds (LOD) value of 6 and

maximum recombination fraction of 0.40. Comparative analyses between the green ash map and maps of two asterid species and two rosid species revealed that synteny has eroded relatively slowly among distantly-related woody Angiosperms.

P0655: Fruit Species

Auxin (IAA) Amino Acid Conjugates Are Hydrolyzed during Diploid Strawberry Seedling Growth

Qian Tang[1], Peng Yu[2], Molly Tillmann[1], Jerry D. Cohen[3] and **Janet P. Slovin**[4], (1)Department of Horticultural Science, University of Minnesota, Saint Paul, MN, (2)Technische Universität München, Freising, Germany, (3)University of Minnesota, St. Paul, MN, (4)USDA/ARS Genetic Improvement of Fruit and Vegetables Lab, Beltsville, MD

The juicy flesh of the strawberry results from enlargement of the stem tip (the receptacle) underlying the carpels in response to auxin and gibberellin produced by the developing achenes, the botanical true fruit. The auxin originates in the achenes, which spiral up the outside of the receptacle. In later stages of berry development, auxin slows the ripening process. The literature describes a pattern of free auxin (IAA) accumulation in the developing berry suggestive of active metabolic and/or transport activity that sustains the enlargement of the receptacle after embryo development is complete. In diploid strawberry, *Fragaria vesca,* embryo development is complete at 10 to 13 days after pollination. Auxin is found in plants primarily conjugated to various amino acids and sugars. Strawberry tissues are capable of synthesizing auxin conjugates, and transcriptome data shows the expression of genes involved in IAA conjugate formation and hydrolysis throughout embryo development and subsequent seedling growth. Using a highly sensitive, high resolution, liquid chromatography-mass spectrometric method, we have now identified all the low molecular weight indole-auxin amino acid conjugates in achenes of *F. vesca* as consisting only of IAA-aspartate and IAA-glutamate. In contrast to what is believed to occur in Arabidopsis, we determined that IAA-aspartate and IAA-glutamate are hydrolyzed by seedlings to provide a source of free IAA for growth.

P0656: Fruit Species

Deciphering the Balance between Sexual and Asexual Plant Reproduction in an Herbaceous Perennial, the Strawberry

Béatrice Denoyes, INRA, Villenave d'Ornon, France

In many polycarpic perennial plants, sexual reproduction is often combined with asexual reproduction; the most common type of asexual reproduction is clonal growth carried out by producing new ramets. In strawberry, flowering and vegetative reproduction constitute highly desirable traits despite they may antagonize each other. Improving productivity in strawberry by controlling the trade-off between sexual reproduction and vegetative propagation is a major challenge.

Perpetual flowering and runnering controls has recently been deciphered in the woody diploid strawberry (*F. vesca*) where , mutations lead to the inactivation of a homologue of the floral repressor *TFL1* and to a *GA20ox* enzyme involved in GA pathway. In the cultivated octoploid strawberry crop, the most consumed berry fruit worldwide, PF trait has been shown to be controlled by a major QTL named *FaPFRU* locus. The *FaPFRU* locus displays opposite effects on flowering (positive effect) and on runnering (negative effect), indicating that both traits share common physiological control. In the diploid, a balance between flowering and runnering is also observed although the flowering and runnering traits are controlled by different loci. The study of the environmental conditions on the balance between flowering and runnerring is today studied in a European project, GoodBerry.

P0657: Fruit Species

A Mapping By Sequencing Approach Identifies a Transposon Insertion in the FvMyb10 Gene As Responsible for White Fruits in F. vesca

Cristina Castillejo[1], Julie Caruana[2], José Vallarino[3], Juan C. Triviño[4], Sonia Osorio[3], José F. Sánchez-Sevilla[1], Zhongchi Liu[2] and **Iraida Amaya**[1], (1)IFAPA Centro de Churriana, Málaga, Spain, (2)University of Maryland, College Park, MD, (3)IHSM-University of Málaga-CSIC, Málaga, Spain, (4)Sistemas Genómicos, Valencia, Spain

Fruit color in the genus *Fragaria* varies widely from completely white fruits to dark red and it is an important trait for strawberry breeders. Gaining insight into the genetic factors affecting natural variation in this trait is key for efficient modification of fruit color in breeding programs.

We applied a NGS-based mapping approach using bulked segregant analysis (BSA) for mapping fruit color in a segregating population derived from a *F. vesca* accession with white fruits. Whole-genome sequencing was performed on two pools of DNA from white- or red-fruited F2 plants. SNPs with different allele frequencies between the pools were identified and plotted along the genome. Significant SNPs were only present in chromosome 1, with the higher ΔSNP-index spanning a region from 11.1 to 18.5 Mb. This interval encompasses ~300 genes containing non-synonymous mutations or frameshifts, including the *FvMyb10* gene. A recent study has shown that white/yellow fruits of several *F. vesca* accessions result from a single nucleotide mutation on *FvMyb10*, but this mutation was not present in the IFAPA white-fruited 596 accession. Instead, we identified a LTR retrotransposon inserted in the third exon of *FvMyb10*. This insertion introduces several in-frame STOP codons which give a truncated protein lacking its C-terminal 141 residues. The presence of the transposon in homozygosis co-segregated with white fruits in the complete F2 mapping population. We are currently performing functional analyses, metabolic profiling and analyzing the presence of this transposon and its possible association to fruit color in other accessions within the genus *Fragaria*.

P0658: Fruit Species

Genome-Wide Association Mapping Uncovers a Dominant Gene Conferring Resistance to Fusarium Wilt in Strawberry

Dominique D.A. Pincot[1], Thomas J. Poorten[1], Julia M. Harshman[1], Charlotte B. Acharya[1], Glenn S. Cole[1], Thomas R. Gordon[2], Michelle Stueven[2], Patrick P. Edger[3] and Steven J. Knapp[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Department of Plant Pathology, University of California, Davis, Davis, CA, (3)Department of Horticulture, Michigan State University, East Lansing, MI

Fusarium wilt, a devastating soil-borne disease caused by the fungal pathogen *Fusarium oxysporum* f. sp. *fragariae*, impacts strawberry (*Fragaria* × *ananassa*) production nearly worldwide. With disruptive changes in soil fumigation practices in recent years, the development and deployment of Fusarium wilt resistant cultivars has become critical. While resistant cultivars have been reported, a limited number of

accessions and pathogen isolates have been analyzed, and contradictory conclusions have been reported in earlier studies to elucidate the genetics of resistance. To identify sources of resistance to Fusarium wilt, clones of 566 historically and commercially important germplasm accessions were artificially inoculated with a virulent isolate of the pathogen and phenotyped in 2016 and 2017. Bimodal phenotypic distributions were observed in both year, and broad-sense heritability across years was 0.96. Loci linked to Fusarium wilt resistance were identified using 38,506 SNPs of the Affymetrix iStraw35 chip and the *Fragaria vesca* 4.0 diploid reference genome. We identified a statistically significant QTL consisting of nine SNPs spanning 2.3Mb on chromosome 2. We predicted the presence of a dominant resistance gene (*FoR2C-1*) on linkage group 2C in the octoploid genome, which was confirmed by genetic mapping in phase-known segregating populations. 93.6% of the accessions had one of two *FoR2C-1* haplotypes, where the resistant allele was present at a frequency equaling 0.16. Using identity-by-descent, the resistant *FoR2C-1* allele was traced to cultivars originating as early as 1935.

P0659: Fruit Species

The Identification of New Molecular Pathways Invoked in the Human Stomach Cancer Cell Lines MKN28 and AGS in Response to Nutraceuticals Isolated from *Rubus* spp.

John David Swanson, Heather J. Axen, Sabrina Fabiano and Julia Gambardella, Salve Regina University, Newport, RI

Gastric adenocarcinoma is the third most deadly cancer, with the five-year survival rate at stage IV being only 4% worldwide. Current treatments include surgery, chemotherapy, and radiation, all of which harm non-cancerous cells. Nutraceuticals, naturally occurring plant compounds, offer a potential area of investigation. Specifically, gallic acid, a phenolic compound found in raspberries and blackberries, has been shown to target cancer cells by halting cell cycle progress. To test the hypothesis that key genes coding for cell cycle progression in cancer are more affected by gallic acid, the changes in gene expression associated with gallic acid treatments in the immortal gastric cancer cell lines MKN-28 and AGS were analyzed through Affymetrix microarrays. MKN-28 and AGS cells were cultured and treated with 0 or 100µM of gallic media and collected at 0, 6, 18, 24, 28, and 36 hours. RNA was extracted, evaluated for quality, then hybridized to microarrays at the Brown University genomics facility. Differential expression of treated versus untreated cancer cells over the time series was determined using Cluster 3.0 and Java Treeview. Heat map analyses showing the regulation of genes suggest that those that play key roles in gastric cancer survival are more affected by the gallic acid treatments.

P0660: Fruit Species

Transcriptome Analysis Identifies Candidate Genes Associated with Epicuticular Wax of Blueberry Fruit

Xinpeng Qi[1], Elizabeth L. Ogden[1], Jose V. Die[2], Mark K. Ehlenfeldt[3], Omar Darwish[4], Nadim Alkharouf[4] and Lisa J. Rowland[1], (1)USDA-ARS Genetic Improvement of Fruits and Vegetables Laboratory, Beltsville, MD, (2)Department of Genetics ETSIAM, University of Córdoba, Córdoba, Spain, (3)USDA-ARS Genetic Improvement of Fruits and Vegetables Laboratory, Chatsworth, NJ, (4)Department of Computer and Information Sciences, Towson University, Towson, MD

Blueberry is of high economic value and one of the richest sources of anthocyanins and antioxidants among berry fruits. Most high quality fresh-market blueberries have an appealing light blue coating or “bloom” due to the presence of a visible heavy epicuticular wax layer. This waxy layer also serves as natural defense against fruit desiccation and deterioration. In this study, we bulked RNA from waxy and non-waxy blueberry progenies from two northern-adapted rabbiteye hybrid populations (‘Nocturne’ x T 300 and ‘Nocturne’ x US 1212), and generated 316.85 million RNA-seq reads. We *de novo* assembled this data set integrated with other publicly available RNA-seq data and trimmed the assembly into a 91,861 blueberry unigene collection. All unigenes were functionally annotated, resulting in 77 genes potentially related to wax accumulation. We compared the expression pattern of waxy and non-waxy progenies using EdgeR and identified overall 1,125 genes in the T 300 population and 2,864 genes in the US 1212 population with at least a two-fold expression difference. Validated by RT-qPCR experiments, two excellent candidate genes have emerged, one with homology to very-long-chain enoyl-CoA reductase and another with homology to acyl-[acyl-carrier-protein] hydrolase, as possibly being responsible for the waxy coating in our populations. Although additional work remains to be done to establish a cause and effect relationship, this study is helping to achieve a greater understanding of epicuticular wax biosynthesis in blueberry. In addition, the blueberry unigene transcriptome collection from this study should facilitate functional annotation of a future chromosomal level blueberry genome.

P0661: Fruit Species

Integrating SNP and Transcriptome Data to Identify Organic Acid-Regulating Loci in Cranberry

Stephanie Fong[1], Joseph Kawash[1], James Polashock[2], Jennifer Johnson-Cicalese[1] and **Nicholi Vorsa**[3], (1)Rutgers University, (2)USDA - ARS, Chatsworth, NJ, (3)Rutgers University, New Brunswick, NJ

American cranberry (Vaccinium macrocarpon) fruit is renowned for its tart flavor. However, due to the high acidity, cranberry products require added sugar to achieve suitable sugar to acid ratios for palatability. Commercial cranberry varieties have an average titratable acidity (TA) of 2.3-3.0% (citric acid equivalents), while other fruits consumed fresh, e.g., apple, have TA below 1%. Two acids contribute largely to cranberry fruit TA: citric acid, imparting a sharp sour taste and malic acid, imparting a mellower acid taste. Cranberry fruit typically has malic and citric acid concentrations of 6-8 mg/g and 8-11 mg/g FW, respectively. A screen of cranberry germplasm identified low citric acid (LCA) and low malic acid (LMA) accessions. Phenotypic segregation indicated both LCA and LMA traits are consistent with recessive single loci with Mendelian inheritance and are independent. Using GBS on three segregating populations (n=118), ~200k potential SNP markers for each low acid trait were identified. RNA-seq performed on bulked (3 low TA and 3 high TA, for each acid) individuals revealed 121 and 92 differentially expressed genes for LMA and LCA phenotypes, respectively, relative to those with ‘normal’ acids. Integrating SNP maps and differential gene expression profiles with phenotype provided an effective means of screening for putative low acid associated markers and candidate genes. One SNP marker was localized to a differentially expressed gene, malate synthase. Further analysis of this SNP marker and others will allow us to implement marker assisted selection and efficient seedling screens for low acid progeny.

P0662: Fruit Species

Building a Vaccinium Community to Advance Blueberry and Cranberry Breeding Programs in US

Massimo Iorizzo, Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC, Rosa Karina Gallardo, School of Economic Sciences, Puyallup Research and Extension Center, Washington State University, St. Puyallup, WA, Patrick Edger, Michigan State University, EAST LANSING, MI, James Polashock, USDA - ARS, Chatsworth, NJ, Cesar Rodriguez-Soana, P. E Marucci Blueberry and Cranberry Research and Extension Center, Rutgers University, Chatsworth, NJ, Juan Zalapa, University of Wisconsin-Madison, Madison, WI, Amaya Atucha, Department of Horticulture, University of Wisconsin-Madison, Madison, WI, Hamid Ashrafi, North Carolina State University, Raleigh, NC, Ebrahiem Babiker, USDA-Agricultural Research Service, Poplarville, MS, Patricio R. Munoz, University of Florida, Gainesville, FL, Nahla Bassil, USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, Mary Ann Lila, Plants for Human Health Institute, Kannapolis, NC, Changying Li, College of Engineering, University of Georgia, Athens, GA, Lara Giongo, Fondazione Edmund Mach, San Michele all'Adige, Italy, Penelope Perkins-Veazie, Professor, Plants for Human Health Institute, North Carolina State Universit, Kannapolis, NC, Nicholi Vorsa, Rutgers University, New Brunswick, NJ and Chad Finn, USDA-ARS, Horticultural Crops Research Unit, Corvallis, OR
This research emerged from a U.S. Department of Agriculture-NIFA funded project to plan efforts to advance the genetic, genomic, and phenotypic resources available for blueberry and cranberry and develop genomic tools to guide and accelerate the development of improved cultivars. To establish blueberry and cranberry industry breeding priorities for fruit and plant quality traits a survey was conducted at 13 commodity group meetings across 12 U.S. states and British Columbia (Canada) between November 2016 and March 2017. Industry responses signaled that the most important trait cluster was fruit quality, particularly fruit firmness, flavor, shelf life, color and size. These fruit quality traits impact the price premium received by producers; influence consumer's preferences; and can facilitate mechanical harvesting and improve processing efficiency, which are critical to the economic viability of the *Vaccinium* industry. Relative importance assigned to traits for disease resistance, arthropod resistance, and tolerance to abiotic stresses differed across regions. These findings are currently used to identify the objectives of a *Vaccinium* multistate and transdisciplinary coordinated project to develop accelerated DNA-based selection strategies to develop cultivars with improved traits for the North American cranberry and blueberry industries.

P0663: Fruit Species

Synthesizing Almond Trees That Produce Peaches using Two Key Fruit Genes

Pere Arús[1], Iban Eduardo[1], Ignasi Batlle[2] and Maria Jose Aranzana[1], (1)IRTA, Barcelona, Spain, (2)IRTA, Constantí, Spain

Almond and peach are two closely related species that can be crossed and produce fertile offspring, but they are distinct in many other aspects: they have been selected under domestication in different environments and for different characters - the peach fleshy fruit and the almond kernel - almond is a self-incompatible, highly variable, adapted to dry and hot conditions, while peach is self-compatible, with a narrow gene pool and requiring optimal growing conditions for production. Almond has been proposed as a source of genes for peach, particularly those related with disease resistance and abiotic stress, which are scarce in the peach commercial germplasm. Two major genes that determine crucial differences between peach and almond fruits were identified and mapped in almond x peach crosses. These genes determine fruit mesocarp thickness (*Alf*) and mesocarp juiciness (*Jui*). Introgressing these two genes into almond could be a good way to engineer plants that have most of the almond variability and produce an edible fruit. The breeding scheme used to produce such synthetic organism, and its possible applications are presented and discussed.

P0664: Fruit Species

Evolutionary Analysis of *Dormancy-Associated Mads-Box* genes in *Prunus*

Kanae Masuda, Hisayo Yamane and Ryutaro Tao, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

Gene duplication positively releases the conflict in redundancy of the gene functions, which has played an important role in the evolution of lineage-specific traits in plants. Six tandemly-repeated *DORMANCY-ASSOCIATED MADS-box* (*DAM*) genes were present in the genome of the genus *Prunus*, such as peach (*P. persica*) and Japanese apricot (*P. mume*), and this duplication seemed to be *Prunus* linage-specific. Here, we identified the *DAM* genes from the genome of sweet cherry (*P. avium*), which is phylogenetically remote from *P. persica* and *P. mume* within *Prunus*, and investigated the evolutionary paths of the establishment of dormancy system by assembling the *Prunus DAM* genes. Phylogenetic approaches defined the duplication patterns across the six *DAMs*, where *DAM4-6* are nested to the same clade. Our expression analysis suggested *P. avium DAM4-6* were up-regulated during dormancy as is the case with other *Prunus* where *DAM4-6* were involved in the regulation of dormancy induction and release. Evolutionary analyses of *Prunus DAM4-6* indicate that *DAM5* had specifically experienced positive selection since the divergence from the other two sisters, on the I-domain, which potentially interacts with counterpart transcription factors. Based on the evolutionary rates estimated, it was suggested that *DAM6* has maintained the original function whereas *DAM4* is released from purifying selective pressure. Taken together, it is suggested that *Prunus* lineage-specific duplications of *DAM* genes could have triggered the evolution of *Prunus*-specific dormancy system.

P0665: Fruit Species

Enhancing Postharvest Tree Fruit Quality with Functional Genomics

Loren A Honaas[1], Heidi Hargarten[1], Joshua P. Der[2], Eric Kenneth Wafula[3], Claude dePamphilis[3], David Rudell[1] and James Mattheis[1], (1)USDA-ARS Tree Fruit Research Lab, Wenatchee, WA, (2)California State University, Fullerton, Fullerton, CA, (3)Penn State University, University Park, PA

Genetic mechanisms that influence pome fruit quality in the postharvest period are poorly understood. As an initial step towards enhanced postharvest fruit management we are exploring fruit transcriptomes to search for genetic factors that influence various aspects of fruit quality. Identification of these genetic factors will build a tool kit for postharvest researchers, and eventually producers, that can be used to finely monitor fruit status during long storage periods. This information might be then used for diagnostics and risk assessment towards enhancing outcomes in the postharvest period. Within proven experimental frameworks we are studying gene expression in tractable cultivar/disorder systems to develop strategies for gene discovery. An impediment to uncovering crucial genetic factors is the need to make discoveries in genetically distinct cultivars (from the Golden Delicious reference genome) where these differences cause a loss of fidelity in gene expression

measurements. Our parallel approach is to develop and refine methods for gene discovery while exploring fruit responses to commercially relevant postharvest treatments and practices.

P0666: Fruit Species

Global Analysis of DNA Methylation Variation Related to Floral Colour Variegation in Peach

Xinxin Wu[1,2], Yong Zhou[1], Zhihong Gao[1] and Zhen Zhang[1], (1)Nanjing Agricultural University, Nanjing, China, (2)Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Nanjing, China

Peach is an important fruit and ornamental plant in the world. Variegation in flowers often receives consumer attention and variegated plants are of high value in the market. To determine the relationship between DNA methylation and phenotype, we obtained the first single-nucleotide resolution DNA methylation of variegation cultivars in peach by bisulfite sequencing (BS-Seq). In our data, a similar methylation rate of 11.96% in red flower buds (RF) and 12.74% in variegated (VF) were determined. The CG methylation is mainly concentrated in transcription regions. We identified 189 differentially methylated regions. Associated with the transcriptional analysis, 8,416 methylated genes showed expression specificity between RF and VF. Methylation level of the coding regions of genes is higher than upstream and downstream, and silent genes have higher methylation levels than expressed genes. Compared with the proteome and transcriptome, 52 different proteins and 106 differently expressed genes have varying degrees of methylation. Among them, the leucoanthocyanidin dioxygenase (*LDOX*) gene related to colour variegation displayed differential methylation. The further showed the expression of *LDOX* gene and the activity of LDOX (EC: 1.14.11.19) enzyme in RF were higher than that of VF and the trend was consisted with the flower colour phenotype. We identified the context and level of methylation at each site and analysed DNA methylation with BS-Seq and found that LDOX might play the key role in the variegated flower petal formation in peach. These results will be valuable for future studies in the variegated flower petal formation in peach.

P0667: Fruit Species

Genome-Wide Analysis of Basic Helix-Loop-Helix (bHLH) Transcription Factors in Peach

Chunhua Zhang, Jiangsu Academy of Agricultural Sciences, Nanjing, China

The bHLH (basic helix-loop-helix) transcription factor family is a superfamily found in all eukaryotes plants, and plays important roles in regulating growth and development. Over the past few decades, many bHLH family genes have been identified and characterized in herbaceous and woody plants, such as in Arabidopsis, rice, apple, and blueberry. However, the bHLH family genes in peaches has not been comprehensively identified and characterized. Here, we identified 115 PpbHLH members in peach (*Prunus persica*) genome, which could be classified into 18 subfamilies according to phylogenetic analysis with bHLH proteins in Arabidopsis. The scaffold location, gene structures and conserved motifs of these PpHLHs were also characterized. Coupled with relative expression analysis of *PpHLH* genes in red-fleshed fruits at four developmental stages of peach, we identified several *PpHLH* genes that might be responsible for anthocyanin biosynthesis. This study provides prerequisite and insight for further studies on the molecular mechanism of these genes involved in regulating red color formation of red-fleshed fruits in peach.

P0668: Fruit Species

Construction of High Density Apricot (*Prunus armeniaca* L.) Linkage Map using SNPs Detected by Genotyping-by-Sequencing (GBS)

Ana Pina[1], Patricia Irisarri[1], Pilar Errea[1] and Tatyana Zhebentyayeva[2], (1)Unidad de Hortofruticultura, Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA), Instituto Agroalimentario de Aragón—IA2 (CITA-Universidad de Zaragoza), Zaragoza, Spain, (2)Clemson University Genomics & Computational Biology Laboratory, Clemson, SC

Several genetic linkage maps were constructed for the apricot (*Prunus armeniaca* L.) genome, using amplified fragment length polymorphisms (AFLPs) and simple sequence repeat (SSR) markers. Genotyping by sequencing (GBS) is a cost-effective alternative for developing thousands SNP markers useful for linkage map construction. In this report we present the first sequence-based genetic map in apricot. A saturated linkage map was developed from an apricot F1 population of 138 individuals, derived from the intra-specific cross of the cultivars 'Moniqui' and 'Paviot' (MxP). Due to contrast parental phenotypes, this cross segregates for several important agronomical characters such as self-compatibility, graft compatibility tendency and fruit color. The parents and individual progeny genotypes were generated at the Apek1 restriction sites. Raw unidirectional reads were processed using Stacks v1.35. Filtered reads were aligned to Peach v2.0 and analysis of distribution along peach scaffolds was performed. Two parental maps as well as consensus map were constructed for downstream applications. The resulting genetic maps composed of eight linkage groups were aligned to other *Prunus* saturated linkage maps. The MxP apricot maps presented here provides a valuable set of sequence-based SNPs useful for identification of quantitative trait loci (QTLs) and further analysis of the genetics of relevant morphological traits in apricot.

P0669: Fruit Species

Genome-Wide Association Study of Partial Resistance to Bacterial Canker in Apricot

Mariem Omrani[1,2], Guillaume Roch[3], Alain Blanc[1], Cindy E. Morris[2] and Jean-Marc Audergon[1], (1)UR1052 GAFL, INRA, Montfavet, France, (2)UR407 Pathologie Végétale, INRA, Montfavet, France, (3)CEP Innovation, Lyon, France

Apricot, a highly valuable crop is threatened by the growing importance of bacterial canker caused by *Pseudomonas syringae*. Among the key factors able to control the disease, genetic improvement is a promising measure. The variability of susceptibility on branches and the characterization of genetic determinants through a genome wide association study were thus investigated on a core-collection.
73 accessions were annually inoculated in the orchard with an aggressive strain of the bacterium for 4 years. Phenotypic data about the length of both external canker (lgc) and superficial browning (bs) of tissues were collected. The analysis displayed a highly environmental-dependent genetic variation with broad-sense heritabilities of lgc and bs reaching respectively 59% and 78% for the most severe year. Considering the two variance-maximizing years, genetic (G) and genetic x year (GxY) BLUP were predicted for each variable using a linear mixed model. Association analysis were performed with a 63,236 SNP set through both a multi-variate (GEMMA) and a multi-locus genome-wide analysis.

By exploiting the between-years (multi-locus model on G and GxY terms) and between-phenotypes (multivariate model on lgc-bs G terms) correlations, 9 significant associations have been detected. Among them, two SNP impacting both lgc and bs expressions over the two studied years and explaining 43% and 33% of the total phenotypic variance were identified on chromosomes 5 and 6. A long-range linkage disequilibrium had been noticed between these two markers suggesting a co-selection effect. The associated SNP reported from this work will open up new opportunities for a Marker-assisted selection strategy.

P0670: Fruit Species

Genotyping by Sequencing and QTL Analysis for Phenolic Compound Content in Japanese Plum (*Prunus salicina* Lindl.)

Igor A. Pacheco Cruz, INTA - Universidad de Chile, Santiago, Chile

Flavonoids are key elements for colour and functional compound content in plant-derived foods; among them, anthocyanins have a bioactivity potential and are key contributors for fruit color in Rosaceae. Discovery of genetic markers for assisted selection of new varieties with increased functional compound content, as well as identification of candidate genes controlling fruit secondary metabolite content, are central objectives in our research. In order to pursue these aims in a commercially important species for Chilean fruit production, we performed an initial genetic analysis for individual anthocyanin content in a Japanese plum (*Prunus salicina* L.) F1 progeny obtained from the cross between selection '98-99' and the cultivar 'Angeleno'. UHPLC-DAD-Orbitrap-MS analyses were employed for compound identification in selected samples. HPLC-DAD was employed for anthocyanin profiling in skin (SK) and flesh (FL) methanolic extracts from fruits harvested at commercial maturity in 90 F1 individuals, for which genotypic data for 4058 SNPs were available after Genotyping-By-Sequencing. In addition, spectrophotometric methods were employed for total phenols, total procyanidins, total flavonoids and antioxidant activity in the F1 family. Two HPLC signals (peaks), previously identified as cyanidin-3-glucoside (C3G) and cyanidin-3-rutinoside (C3R) were detected in concentrations hundreds to thousand times higher than the other detected peaks. C3G and C3R, as well as minor-content anthocyanins (here called "M" and "K"; identification in progress) showed significant genetic contributions to phenotypic variance ($p < 0{,}001$). Parametric and non-parametric QTL-analyses (Interval Mapping and Kruskal-Wallis, respectively) were performed using linkage maps for both parents, separately. Significant marker-trait associations ($p < 0{,}05$) were detected mainly for compounds present in skin samples. QTLs located in linkage-groups (LG) 1, 3, 4, 5 and 7 of *Prunus* genome explained between 38% and 45% of the total phenotypic variance. The results here presented constitute the first elements in a study for genetic architecture dissection of potentially bioactive secondary metabolite content in Japanese plum fruits. This work is being supported by grants FONDECYT inicio 11150662, PAI-CONICYT 79140020, FONDECYT postdoctorate 3160080 and FONDAP Center for Genomic Regulation.

P0671: Fruit Species

Comparative Genomics for Three Sweet Cherry Varieties

Javier Cáceres-Molina[1], Tomás Carrasco-Valenzuela[1], Claudio Urra[1], **Karin Rothkegel**[1], Humberto Prieto[2], Claudio Meneses[1,3] and Andrea Miyasaka de Almeida[1], (1)Universidad Andrés Bello, Centro de Biotecnología Vegetal, Santiago, Chile, (2)Instituto de Investigaciones Agropecuarias - La Platina Station, Santiago, Chile, (3)FONDAP, Center for Genome Regulation, Santiago, Chile

Sweet cherry (*Prunus avium*, 2n=2x=16) belongs to the *Rosaceae* family and is one of the most produced and exported fresh fruit from Chile, the main producer of the Southern Hemisphere. Developing of sweet cherry varieties well adapted to local conditions, with high crop yields represents a major challenge and genomic-based breeding programs could help to address it. Despite its commercial importance, there's a lack of genomic information in order to facilitate the identification of features associated to agronomical important traits. Our proposal was to perform an *in silico* comparative genomics strategy to search for meaningful structural differences between three sweet cherry varieties 'Karina', 'Kordia' and 'Royal Dawn', comparing them to the recently published genome of the Japanese variety 'Satonishiki'. A whole genome sequencing strategy using Next-Generation Sequencing technologies was performed, obtaining paired-end libraries with a total of 434.6 millions of reads for 'Karina', 214.2 millions for Kordia' and 273.6 millions for the 'Royal Dawn' variety. These reads were mapped to the genomic sequences of 'Satonishiki' with a mapping efficiency of 90.5%, 87.9% and 86.6% respectively. Furthermore, we have successfully mapped 7078 SNP publically available to the reference genomic sequences in order to guide the variant discovery. Finally, using the mapped reads, we have constructed a partial draft for each of the three varieties (N50 ~0.22 Mbp and longest scaffold ~1.4 Mb) for using in further genomic studies.

Funding: FONDEF G09I1008 and CORFO 13CTI250-SP05.

P0672: Fruit Species

NGS Analysis Reveals a Possible Causal Mutation Conferring Self-Compatibility in a Sweet Cherry Cultivar Cristobalina

Kentaro Ono[1], Takuya Morimoto[1], Takashi Akagi[1,2], Ana Wünsch[3] and **Ryutaro Tao**[1], (1)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (2)JST-PRESTO, Saitama, Japan, (3)CITA de Aragón,, Zaragosa, Spain

Self-incompatibility (SI) is an important reproductive mechanism to maintain genetic diversity within a plant species. However, SI can be a limiting factor for efficient production and breeding of cultivated plants. *Prunus* exhibits the S-RNase-based gametophytic SI system. Many self-compatible (SC) mutants of *Prunus* fruit tree species were found during a long history of the cultivation and have been utilized extensively for commercial production. 'Cristobalina', an SC sweet cherry (*Prunus avium*) cultivar, is one such example and presumed to have arisen by a mutation of a pollen-part modifier gene that is located on LG3, outside of the *S* locus. In this study, we conducted a subsequence analysis on whole-genome sequences obtained by Illumina sequencing to identify a causal mutation in 'Cristobalina'. Two 'Cristobalina' F_1 populations, both segregating for SC and SI individuals, were subjected to Illumina sequencing. Obtained reads were subdivided into 35-bp subsequences called k-mers. K-mers thus obtained were cataloged into SC and SI pools, and SC-specific k-mers were extracted. Then, the original reads containing the SC-specific k-mers were assembled into candidate contigs containing SC locus of 'Cristobalina'. Next, we further checked SC-specificity of the contigs utilizing Illumina genomic reads from various sweet cherry cultivars and 'Cristobalina' progenies. Comparisons of the SC-specific genomic contigs obtained and pollen mRNA-Seq data revealed a possible causal mutation for SC in 'Cristobalina'. Functional characterization of the modifier gene would lead to further understandings of the SI reaction in S-RNase-based gametophytic SI system in *Prunus*.

P0673: Fruit Species

Comparative Analysis of Cytokinin Modulated Genes in SWEET Cherry (Prunus avium) and Peach (Prunus persica) Fruits

Karen Mujica, Claudio Miguel Ponce and **Lee Meisel**, Instituto de Nutrición y Tecnología de los Alimentos (INTA), Universidad de Chile, Macul, Chile

The non-climateric fruit, sweet cherry and the climateric fruit, peach, are Chilean high yield export crops. The international demand for these stone fruits, fresh and processed, is due to the nutritional and bioactive/medicinal properties that they contain. These desirable traits, formed during fruit maturity/ripening, are mediated by phytohormones, such as cytokinin. However, the molecular mechanisms involved in these phenological traits are not well documented. A better understanding of the underlying mechanisms that modulate the ripening process in stone fruits needs to be obtained to improve these desirable traits. Recently, meta-analyses in *Arabidopsis* have revealed a conserved group of cytokinin-induced genes. In order to further understand the role of cytokinin in fruit maturity/ripening, the objective of this work was to identify cytokinin-responsive genes in sweet cherry and peach fruits during the maturity/ripening process. Using bidirectional blasts we have identified many putative orthologs of the cytokinin-induced genes in sweet cherries and peaches. Gene expression analyses have revealed that there is a conserved cytokinin responsiveness to 15 of these orthologs in both cherry and peaches. Comparative analyses, of the sweet cherry and peach genomic *loci* of these orthologs, reveal conservation and divergences of several known ARR type-B response elements. By further studying these conserved and divergent *loci*, we may begin to better understand the role of cytokinin in fruit maturity/ripening and how this role has diversified between climacteric and non-climateric fruits.

This work was financed by *CONICYT Fondecyt Regular N°1171016*; *CONICYT Fondecyt Regular N°1121021*; *ENLACE-VID -Universidad de CHILE ENL006/2016*; *Beca Doctorado Nacional #21140426*

P0674: Fruit Species

Genetic Determinism of Flowering and Maturity Dates in Sweet Cherry (Prunus avium)

Jose Quero-Garcia[1], José Antonio Campoy[2], Teresa Barreneche[1], Bénédicte Wenden[1], Mathieu Fouché[1], Loïck Le Dantec[1], Gregorio Lopez-Ortega[3], Anton Ivancic[4], Daniela Giovannini[5], Felicidad Fernandez[6] and Elisabeth Dirlewanger[1], (1)INRA, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France, (2)CEBAS CSIC, Espinardo, Spain, (3)IMIDA, Alberca las Torres, Spain, (4)University of Maribor, Hoce, Slovenia, (5)CREA, Forli, Italy, (6)East Malling Research, East Malling, Kent, United Kingdom

Sweet cherry is an economically important fruit species which is seriously threatened by the impact of climate change. Indeed, an increase in autumn and winter temperatures will impair the proper satisfaction of chilling requirements for flowering, with a subsequent loss of productivity during years of warm winters. Furthermore, an increase in spring temperature will lead to an advance in flowering and an increased risk of frost damage. Hence, breeders search for better adapted varieties in terms of flowering date, and its components chilling and heat requirements. Maturity period is also a critical trait for cherry growers, who seek new varieties enlarging the harvest season, which is particularly short for this fruit crop. The study of the genetic determinism of these traits was initiated at INRA-Bordeaux (France) more than 10 years ago. In this work, we present an update of the QTL analyses conducted over three mapping progenies ('Regina' × 'Lapins', 'Regina' × 'Garnet' and 'Fercer' × 'X', composed of 122, 117 and 67 individuals, respectively), with a range of 6 to 11 years of data for traits flowering and maturity dates. Two years of data on chilling and heat requirements evaluated on progeny 'Regina' × 'Lapins', as well as the first year of data for flowering and maturity dates from a European multi-site trial (6 sites across 5 different countries, France, Slovenia, Spain, Italy and England) involving progeny 'Regina' × 'Lapins', will be described.

P0675: Fruit Species

Last Advances on Sweet Cherry Genomic Tools for the Pre-Breeding

Mathieu Fouché[1], Nathalia Helena Zaracho Echagüe[1], Teresa Barreneche[1], Laurent Richard[1], Noémie Vimont[2], Bénédicte Wenden[1], Timothee Flutre[3], Loïck Le Dantec[1], Jose Quero-Garcia[1] and Elisabeth Dirlewanger[1], (1)INRA, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France, (2)INRA-UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France, (3)INRA UMR 1334 AGAP Amélioration génétique et adaptation des plantes méditerranéennes, Montpellier, France

Breeding of perennial trees such as sweet cherry is a long and expensive process. Indeed, due to a long juvenile phase, it can take more than 20 years to create a new variety. In face of the economic importance of sweet cherry on the fruit production and to answer more quickly to the future challenges an efficient breeding program is essential. One of our team's goal is to understand sweet cherry adaptive responses to climate change in order to create new sweet cherry varieties well adapted to the global warming, with good yield and fruit quality. We are focused on traits such as chilling and heat requirements for flowering, which allow a correct flowering at the right time, as well as the fruit weight, firmness and additional fruit quality traits in order to meet farmer's needs. To optimize the breeding process, we generated genomics tools such as a high quality genome sequence, high density genetic maps, QTL detection, RNAseq data that contribute to the development of molecular markers and the identification of candidate genes to perform Molecular Assisted Selection (MAS) on some traits of interest. Moreover, we are currently testing the advantages offered by the genomic selection methodologies to select our hybrids for highly complex traits, both in terms of genetic determinism (highly polygenic) and in terms of complexity of phenotyping. Altogether, these genomics tools will considerably decrease the cost and the duration (higher number of hybrids could be screened in a year) of our sweet cherry breeding program. Here, we present some of these tools and examples of molecular markers usable in MAS.

P0676: Fruit Species

A High Quality Draft Assembly of the Diploid 'Regina' Sweet Cherry Genome

Loïck Le Dantec[1], Nabil Girollet[2], **Mathieu Fouché**[1], Jérôme Gouzy[3], Erika Sallet[3], Jose Quero-Garcia[1], Teresa Barreneche[1], Bénédicte Wenden[1] and Elisabeth Dirlewanger[1], (1)INRA, UMR 1332 Biologie du Fruit et Pathologie, Villenave d'Ornon, France,

(2)UMR EGFV 1287 - Ecophysiologie et Génomique Fonctionnelle de la Vigne, Villenave d'Ornon, France, (3)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France
Sweet cherry (*Prunus avium* L.) is a diploid species with an estimated genome size of 338 MB (Arumuganathan *et al.* 1991). It is mostly self-incompatible, and therefore has a heterozygous genetic background. We have sequenced and assembled the genome of the 'Regina' sweet cherry variety, which is a late blooming cultivar. Here we present a draft phased sweet cherry genome assembly using a combination of sequencing strategies (long reads sequencing with PacBio RSII and optical mapping with BioNano whole-genome maps). PacBio RSII (82X) long reads were de novo assembled using FALCON assembler and phased using FALCON UNZIP. The assembly was polished using Quiver and BioNano optical whole genome maps were used for further scaffolding. Our de novo genome assembly resulted in a genome of 279 Mb (83 % of estimated genome size), an N50 of 5,96 MB, 92 scaffolds and a largest scaffold of 16,3 Mb. We are now in the process of reconstructing the pseudo molecules using high density genetic linkage maps and GBS data. Annotation will be done with the integrative EuGene platform using different sources of evidence available for gene prediction including transcriptomic data from Regina obtained previously in the lab (Unigene set and RNASeq).

P0677: Fruit Species

Genomewide Selection and Germplasm Architecture in Apple (Malus domestica)

Elizabeth Nicole Blissett[1], Steven McKay[2], Nicholas Howard[1], James J. Luby[3] and Rex Bernardo[3], (1)University of Minnesota, St Paul, MN, (2)DuPont Pioneer, Johnston, IA, (3)University of Minnesota, St. Paul, MN
Apple (*Malus domestica*) breeders need to develop cultivars that meet the demands of consumers, growers, and distributors. New cultivars must meet or exceed standards for many traits such as flavor, texture, appearance, size, and storability. Many of these traits are controlled by a large number of small-effect quantitative trait loci (QTL). However, current genetic tests in apple only allow for selection on a few traits which are controlled by major-effect QTL. Genomewide selection is already widely used in row crops, but has been more slowly adopted in highly heterozygous, perennial, asexually propagated crops like apple. Our objectives were to assess whether accounting for heterozygosity or treating major QTL as having fixed effects improves the accuracy of prediction. Our results on sensory traits in a set of germplasm indicated the potential for use of genomewide selection in the University of Minnesota breeding program. To confirm this, additional analyses with the same prediction models are being conducted in a more targeted set of germplasm. Additionally, there is little knowledge of the distribution of favorable alleles, called germplasm architecture, for important traits across diverse apple germplasm. Understanding germplasm architecture would allow breeders to make more informed parental selections. Both improved genomewide selection models and knowledge of germplasm architecture offer the potential to increase the efficiency of apple breeding.

P0678: Fruit Species

Fox Hunting in Wild Apple: searching for Novel Apple Genes in *Malus sieversii*

Michael Wisniewski[1], John L. Norelli[1], Timothy Artlip[2], Erik Burchard[2] and Jing Ma[1], (1)USDA-ARS, Kearneysville, WV, (2)USDA-ARS-AFRS, Kearneysville, WV
M. sieversii represents a source of genetic diversity for economically-important apple traits, including stress and disease resistance, as well as unique fruit quality attributes. USDA-sponsored expeditions to Central Asia have resulted in a large collection of *M. sieversii* accessions that are maintained at the USDA Plant Genetics Resources Unit (PGRU) in Geneva, NY. *M. sieversii* - PI 613981, is one of the elite lines that was originally selected for collection in Kazakhstan as budwood for its potential drought tolerance and disease resistance. The identification of genes responsible for these and other traits, in *M. sieversii*, however, have not been explored. The characterization of functional genes in wild apple and the establishment of independent overexpressing lines in Arabidopsis that could be used for high-throughput screening would represent a valuable resource for studying apple traits. Thus far, large-scale, high-throughput screening for stress tolerance genes in apple has not been conducted. In the present project, the FOX (Full-length cDNA Over-eXpressing) gene hunting system, which represents an alternative gain-of-function gene hunting technique, has been used to generate 10 - 12,000 gain-of-function mutant lines in *Arabidopsis* carrying independent apple cDNAs derived from a cDNA library of *M. sieversii* – PI613981 constructed from mid-winter bark tissues. The high-throughput screening effort will focus on identifying genes related to freezing tolerance, salt tolerance, drought tolerance, and morphological traits (dwarfing, early-flowering, branching, root architecture, etc.). Initial characterization of the system, a sampling of inserted apple genes, and examples of morphological mutants will be presented.

P0679: Fruit Species

Function and Regulation of *MIR168a* and Its Target *ARGONAUTE1* in *Malus Hupehensis* Defense Responses Against *Botryosphaeria Dothidea*

Xinyi Yu and Shenchun Qu, Nanjing Agricultural University, Nanjing, China
MicroRNA (miRNA)-mediated post-transcriptional regulation plays a fundamental role in various plant physiological processes, including responses to pathogens. MicroRNA168 has been implicated as an essential factor of miRNA pathways by targeting *ARGONAUTE1* (*AGO1*), the core component of the RNA-induced silencing complex (RISC). A fluctuation in *AGO1* expression influences various plant-pathogen interactions, and the homeostasis of *AGO1* and miR168 accumulation is maintained by a complicated feedback regulatory loop. In this study, the connection between miR168 and the resistance of *Malus hupehensis* to *Botryosphaeria dothidea* is revealed. The induction of both the mature miR168 and its precursor in plants subjected to B.dothidea infection indicate the transcriptional activation of *MIR168a*. *MIR168a* promoter analysis demonstrates that the promoter can be activated by *B. dothidea* and salicylic acid (SA). However, the direct target of miR168, *M. hupehensis ARGONAUTE1* (*MhAGO1*), is shown to be induced under the infection. Expression and transcription activity analysis demonstrate the transcriptional activation and the post-transcriptional suppression of *MhAGO1* in response to *B. dothidea* infection. By inhibiting reactive oxygen species (ROS) production and enhancing SA-mediated defense responses, miR168a delays the symptom development of leaves inoculated with *B. dothidea* and impedes the pathogen growth, while *MhAGO1* is found to have the opposite effects. Collectively, these findings suggest that the expression of miR168 and *MhAGO1* in *M. hupehensis* in response to *B. dothidea* infection is

regulated by a complicated mechanism. Targeting to *MhAGO1*, a negative regulator, miR168 plays a positive role in the resistance by alterations in diverse defense responses.

P0680: Fruit Species

Transcriptome of Wax Apple (Syzygium samarangense) Provides Insights into Nitric Oxide-Induced Delays of Postharvest Cottony Softening

Fahe Chen, Jimei University, Xiamen, China

Wax apple (*Syzygium samarangense*) is an important tropical fruit crop in Southeast Asia. The rapid cottony softening that occurs after harvest significantly influences the nutrition, flavour and marker value of the fruit. Nitric oxide (NO) fumigation has been used to delay the cottony softening process in wax apple; however, the underlying molecular mechanisms at the gene regulation level are poorly understood.

To assess gene expression in the wax apple fruit, 134,199 transcripts with an N50 length of 2,447 bp were assembled with the Illumina HiSeq 2000 platform. Using transcriptome annotations, the gene expression between the NO treatment (10 $\mu L \bullet L^{-1}$ NO) and control groups was compared at days 0, 2, 4, 8 and 12 postharvest. The results indicated that important genes in the reactive oxygen metabolism, carbohydrate metabolism and phenylalanine metabolism pathways were significantly regulated by NO gas, which suggests that homeostasis of reactive oxygen species (ROS) and cell wall component biosynthesis may play a central role in NO-induced cottony softening delays.

This study developed transcriptome data resources for wax apple fruit and provided a foundation for understanding the molecular mechanisms underlying NO-induced cottony softening delay. The dynamic analysis of the gene expression patterns suggested that NO-induced delays of cottony softening occur in wax apple via the regulation of cell wall degradation, carbohydrate metabolism, oxidation-reduction and plant hormone signal transduction pathways. These results provide a reference for the study of complicated metabolism in non-model perennial species.

Keywords: wax apple; transcriptome; cottony softening; nitric oxide; gene expression

P0681: Fruit Species

Color Mutations of Fuji Apple (Malus domestica Borhk cv. Fuji) and Their Molecular Mechanisms

Yongbing Yuan, Qingdao Agricultural University, Qingdao, China

Anthocyanins and carotenoids are the most important pigments for red skin and yellow flesh respectively in apples (*Malus domestica* Borhk). 'Yanfu 0', 'Yanfu 1 ',' Yanfu 3 ',' Yanfu 8 ',' Yanfu 9 ' and 'Shoufu ' are a series of deep red skin mutant and 'Beni Shogun' is a early maturity and yellow flesh mutant of 'Fuji' apple. The study on the molecular mechanisms of these colorful mutation showed that among six genes expression of *MdUVR8* , *MdCRY1*, *MdHY5*, *MdMYB1*, *MdDFR*(dihydroflavonol 4-reductase), *MdUFGT* (UDP-glucose: flavonoid 3-O-glucosyltransferase) analyzed using qRT-PCR, *MdUFGT* and *MdHY5* gene expression were found higher significantly in all red skin mutant than control. It is postulated that *MdUFGT* and *MdHY5* gene played important regulatory roles in the anthocyanin biosynthesis of deep red skin mutant of 'Fuji' apples. To analysis the regulation of carotenoid biosynthesis in 'Beni' apple, *MdAP2* (KC415239) belonging to *APETALA2* (*AP2*) family and *MdMADS8* (KC415237) to *SEPALLATA* (*SEP*) subfamily were isolated, and the two transcript factors and several genes which related with carotenoid synthesis were measured by quantitative real time PCR (qRT-PCR) and Pearson correlation (r) analysis in flesh and skin of 'Beni' and 'Kiku'. *MdAP2* and *MdMADS8* were transformed into *Nicotiana benthamiana* and identified the relation genes expression. The results showed that *MdAP2* had positive effect on *PSY* and *LCY-β* expression, *MdMADS8* played an important role in *ZDS* expression. It is postulated that the two transcript factors may play critical roles in carotene accumulation through influencing the expression levels of key carotenoid synthesis genes.

P0682: Fruit Species

Development of Distinguishable Markers Among Apple Sports (Malus × domestica Borkh.) from 'Fuji' and 'Hongro' Using Sequence-Specific Amplified Polymorphism Markers

Hyeondae Han[1,2], Keumsun Kim[1,2], Sewon Oh[1,2], Min-Gi Lee[1,2], Kidong Hwang[1,2], Suji Ahn[2], Sunheum Cho[1,2] and Daeil Kim[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea

Retrotransposons insert large number of copies into the genome, which change the size and generate insertion mutation, thus retrotransposon-based molecular markers could be a useful tool for genetic diversity analysis. The purpose of this research is to distinguish the 'Fuji' sports using sequence-specific amplification polymorphism (S-SAP) markers. Here, we detected two retrotransposons in apple genome sequences. To investigate the potential utility of the two retrotransposons as molecular markers, the 32 primer combinations including 2 long terminal repeat primers and 16 adaptor selective primers were used to develop the S-SAP marker system. Multiple polymorphisms were analyzed by capillary electrophoresis and 5 'Fuji' sport accessions including 'Fuji KIKU 8', 'Yataka Fuji', 'Fidex', 'Jubilee Fuji', and 'Highland Fuji' were discriminated by polymorphic band. In this study, we have successfully developed S-SAP markers and our retrotransposon based S-SAP markers were effective at identifying 'Fuji' sports. It is anticipated that S-SAP markers could be applied in apple genetic research and breeding.

P0683: Fruit Species

Parentage Analysis of Apple Cultivars (*Malus* × *domestica* Borkh.) using SNPs Derived from Genotyping-by-Sequencing

Heon Kyu LIM[1], Min-Gi Lee[1,2], Keumsun Kim[1,2], Sewon Oh[1,2], Hyeondae Han[1,2], Kidong Hwang[1,2], Sunheum Cho[1,2] and Daeil Kim[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea

In our previous study, thirty-three apple cultivars were distinguished by the cleaved amplified polymorphic sequence markers derived from single nucleotide polymorphisms (SNPs) based genotyping-by sequencing. This study analyzed genetic relationship using previously identified SNPs for constructing phylogenetic tree. The grouping was mostly corresponded to their parent cultivars and bud sports, however, 'Sobaek No. 2' (sports of 'Hongro') and 'Sobaek No. 3' (sports of 'Fuji') were not clustered together with their parent cultivars. To identify genetic relationship of those two species, paternity analysis using CERVUS 3.0.7 software and their self-incompatibility relationship analysis were

performed. These results show that 'Jonathan' is the male parent of 'Sobaek No. 2' and 'Sobaek No. 3'. This study infers that the actual parentage of 'Sobaek No. 2' and 'Sobaek No. 3' differed from their reported parentage and will help to better understand the parentage and origins of the apple cultivars.

P0684: Fruit Species

Identifying the Genetic Basis of Resistance to Apple Canker, *Neonectria ditissima*

Amanda Karlström[1], Antonio Gómez-Cortecero[1], Javier Pinedo-Gomez[1], Jim Dunwell[2], Matthew Ordidge[2], Richard J. Harrison[1] and Charlotte F Nellist[1], (1)NIAB EMR, East Malling, United Kingdom, (2)The University of Reading, UK, READING, United Kingdom

Apple canker, caused by the fungal pathogen *Neonectria ditissima,* is an economically damaging disease in apple producing regions of the world – especially in areas with moderate temperatures and high rainfall. The pathogen affects a wide range of hard-wood perennial species causing trunk cankers, dieback and branch lesions in its hosts. Although varieties tolerant to the disease are the most effective method of controlling apple canker, little is known of the genetic basis for the quantitative resistance.

The aim of our project is to use a multiparental population of apple to map quantitative trait loci (QTL) for canker resistance through a pedigree-based analysis. The population, which consists of five full-sibling crosses with a connected pedigree, will be genotyped using the Illumina Inifinium 20k SNP array. Several methods will then be utilised to gather robust phenotypic data on the resistance response of the population. To further narrow down which genes within the QTL are involved in tolerance to *N. ditissima,* resistance gene enrichment sequencing (RenSeq) will be carried out on the parental genotypes.

To date, we have screened the multiparental population for disease response by artificially inoculating detached dormant shoots with a single isolate of *N. ditissima*. Disease progression was measured as the length of developed lesions at regular intervals. Parents and progenitors of the population were screened during the same event. The distribution of the disease response was normal in each of the five seedling families – supporting the premise that resistance to *N. ditissima* is quantitatively controlled in apple. None of the progeny exhibited full resistance, which is consistent with previous findings.

P0685: Fruit Species

Genetic Diversity and Population Structure of Pear Germplasms (*Pyrus* spp.) using Genotyping-by-Sequencing

Keumsun Kim[1,2], Sewon Oh[1,2], Hyeondae Han[1,2], Kidong Hwang[1,2], Min-Gi Lee[1,2], Suji Ahn[1,2], Sunheum Cho[1,2], Yoon-Kyeong Kim[3] and Daeil Kim[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea, (3)National Institute of Horticultural & Herbal Science, Naju, South Korea

Estimation of genetic diversity among *Pyrus* spp. is often very difficult owing to the low morphological diversity, lack of differentiating characters among species, and widespread crossability. In present study, we conducted genotyping-by-sequencing technology to estimate the genetically diverse collection of 231 pear accessions and detected 10,186 SNPs. Phylogenetic tree was constructed using MEGA6 with neighbor-joining method, population structure was estimated by using STRUCTURE v.2.3.4., and principal component analysis (PCA) was performed using TASSEL v5.2.39 software based on the set of 10,186 SNPs. Analysis of phylogenetic tree and genetic structure substantiate the identification of tree distinct subpopulations. The first subpopulation contains mainly *P. communis*, 4 primary pears (*P. glabra*, *P. amygdaliformis*, *P. nivallis*, and *P. elaeagrifolia*), and hybrids between *P. pyrifolia* and *P. communis*. The second subpopulation includes Japanese pears (*P. pyrifolia*) and the third subpopulation comprises Chinese pears (*P. bretschneideri*, *P. ussuriensis*, and *P. betulifolia*) and Korean native pears. Also, PCA was clearly separated into three groups across the first two axes and consistent with phylogenetic tree. Genetic analysis within each subpopulation revealed patterns of diversity associated with geographical origin. Finally, our findings could be useful for developing a core collection of pear genetic data.

P0686: Fruit Species

Transcriptome and Metabolite Profiling Reveals That Regulation of Cuticular Wax Formation during Fruit Development of 'Yuluxiang' Pear

Xiao Wu, Hao Yin and Shaoling Zhang, Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, China

Fruit cuticular wax layer acts as the first protective barrier against fruit splitting and plays pivotal roles in the reduction of pathogenic and insect attacks, protection against mechanical damage, and extension of the storage period. However, the key factors and mechanisms regarding the wax biosynthesis in pear fruit are still unclear. In this study, cuticular wax of 'Yuluxiang' pear fruit during development were analyzed from the perspectives of morphology, transcription and metabolomics. The results demonstrated wax plates were accumulated and their structures were changed during fruit development. Wax content during fruit development was first increased from 0.64 to 1.75 mg/cm^2 between 20 and 80 DAFB and then decreased from 1.75 to 1.22 mg/cm^2 between 80 and 140 DAFB. Transcriptome and metabolite profiling results showed the different accumulation patterns of wax and their key time points respectively during 'Yuluxiang' pear fruit development. The key genes involved in synthesis, transport and regulation of pear fruit cuticular wax were then predicted, which would facilitate their further functional characterization. In addition, the key regulator factors of cuticular wax formation were explored. Based on these results, this study would provide an important back-ground for future studies of pear fruit cuticular wax. Besides, our data also provide important information for exploring the association between wax and the general regulation mechanisms of their formation in plants.

P0687: Fruit Species

Molecular Characterization of the SPL Gene Family in Pear

Peng Cao, Peng Wang, Shaoling Zhang and Juyou Wu, Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, China

Abstract: SQUAMOSA promoter binding protein-like□*SPL*□encodes plants-specific transcription factors playing vital regulatory roles in plant growth and development. There is no information about *SPLs* in 'Dangshan suli' (Chinese white pear), a significant fruit worldwide. Comparative analysis showed sequence conservation between *PbSPLs* and their Arabidopsis counterparts. A phylogenetic treclusters *PbSPLs* into five groups. Seventeen *SPL* genes were as certained to contain the putative miR156 binding site, with 12 and 5 of the genes targeted by miR156 at the coding and 3′UTR region, respectively. *PbSPLs* were differentially expressed in various tissues of Pear. Overexpression *PbSPL15* gene decreased number of the rosette leaves and early the flowering time than wild type.

P0688: Fruit Species

The Unique Evolutionary Pattern of the Hydroxyproline-Rich Glycoproteins Superfamily in the Chinese White Pear (Pyrus bretschneideri)

Huijun Jiao, Nanjing Agricultural University, Nanjing, China

The hydroxyproline-rich glycoprotein (HRGP) superfamily, comprising three families (arabinogalactan-proteins, AGPs; extensins, EXTs; and proline-rich proteins, PRPs), is a class of proline-rich proteins that exhibit high diversity and are involved in many aspects of plant biology. In this study, 838 HRGPs were identified from the pear (*Pyrus bretschneideri*) by searching for biased amino acid composition and conserved motifs. Whole genome duplication (WGD) and dispersed duplication are the major forces driving *HRGP* expansion. The recent WGD event shared by apple and pear generated most of the duplicated HRGPs in the pear. This duplication event drives the variation of the *HRGP* gene structures encoding hydroxyproline (Hyp)-rich motifs. The rate of *HRGP* evolution mainly impacted the Hyp-rich motifs even in chimeric HRGPs. During the evolution of 53 *PRPs* that are also typified by 7-deoxyloganetin glucosyltransferase-like genes, the duplication from *PRP* to *non-PRP* was indirectly modified by positive selection. During the evolution of *AGPs* classified into Pollen Ole e I family, the mutation from *AGP* to *non-AGP* was directly influenced by positive selection. The expression divergence of *HRGPs* was higher than that of other commonly duplicated genes. In the pear pistil, 601 *HRGPs* exhibited expression, while in the pear pollen, 285 *HRGPs* were observed. The qPCR results revealed that *Pbr036330.1* and *Pbr010506.1* are differential expressed for self-incompatibility in pear pistil. The results indicate that the highly variable Hyp-motifs affect the expansion, evolution and expression divergence of HRGPs and that this divergence is responsible for the gain of new functions in plants.

P0689: Fruit Species

PbCOL8 Is a Clock-Regulated Flowering Time Repressor in Pear

Zhe Liu, Nanjing Agricultural University, Nanjing, China

The floral transition is controlled by diverse endogenous and exogenous cues. In many species, COL(CONSTANS-like) genes integrate light and circadian clock signals to regulate flowering time. However, little is known about COLs in perennial woody plants. Here, we identified 15 PbCOLs in pear (Pyrus bretschneideri). PbCOLs were classified into three groups by phylogenetic tree analysis using protein sequences. Multiple sequence alignment analysis revealed conserved B-box and CCT (CO, CO-like, and TOC1) domains in all PbCOL members. This result suggested that PbCOLs might possess conserved functions as other species. Six PbCOLs were found to be regulated by both circadian clock and photoperiod. Here, we showed that PbCOL8, a member of group 2, suppressed the flowering signal integrators FT and SOC1 and could repress flowering time. These findings will contribute to elucidation of the mechanism of floral initiation in pear.

P0690: Fruit Species

Candidate QTL Regions Associated with Fruit Size and Weight in Mapping Population between 'Whangkeumbae' and 'Minibae' (Pyrus spp.)

Hyeondae Han[1,2], Keumsun Kim[1,2], Sewon Oh[1,2], Min-Gi Lee[1,2], Kidong Hwang[1,2], Suji Ahn[1,2], Sunheum Cho[1,2], Yoon-Kyeong Kim[3] and Daeil Kim[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea, (3)National Institute of Horticultural & Herbal Science, Naju, South Korea

Quantitative trait loci (QTL) analysis was performed to identify the QTLs associated with fruit weight and vertical and transverse diameter in the mapping population of 'Whangkeumbae' and 'Minibae'. Five fruits per each individual were harvested at their respective fruit ripening stage and afterwards weight and vertical and transverse diameter were evaluated for 178 F_1 progenies and their parents in 2013, 2014, and 2016. Phenotypic data combined with genotypic data derived from JoinMap 4.1 were loaded into MapQTL 5.0 for QTL analysis. Interval mapping was applied to estimate the candidate regions followed by permutation test with 1,000 replicates. High correlation was observed between fruit size and fruit weight. Although slight differences were observed for fruit traits between years, fruit traits showed normal distribution as with distinctness of parents. Most candidate QTLs for weight, vertical, and transverse diameter were commonly located at linkage group (LG) 1, 9, 10, and 14. Some QTLs specific to vertical and transverse diameter were observed in LG 11 and 15, respectively. The results of study should be validated for identification of coding regions associated with respective target traits using genome annotation, and would be useful information for pear breeding and genetics.

P0691: Fruit Species

Construction of a High-Density Genetic Linkage Map Using Various Codominant Markers in Interspecific Asian Pears (*Pyrus spp.*)

Keumsun Kim[1,2], Hyeondae Han[1,2], Sewon Oh[1,2], Min-Gi Lee[1,2], Kidong Hwang[1,2], Suji Ahn[1,2], Sunheum Cho[1,2], Yoon-Kyeong Kim[3] and Daeil Kim[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea, (3)National Institute of Horticultural & Herbal Science, Naju, South Korea

A high-density linkage map is the first prerequisite for analysis of quantitative trait loci, positional cloning, marker-assisted selection, and comparative mapping in pears (*Pyrus* spp.). This study was conducted to construct an integrated genetic linkage map using the co-dominant markers developed by next generation sequencing technologies. A total of 93 F_1 individuals derived from a cross of 'Whangkeumbae' ×

'Minibae', were used to construct genetic linkage map using single nucleotide polymorphism (SNP), insertion/deletion (InDel), and simple sequence repeat (SSR) markers. Overall, 2,361 of codominant molecular markers including 2,331 SNPs, 13 InDels, and 17 SSRs were loaded to JoinMap 4.1 for linkage analysis. The genetic linkage map contained 1,488 markers in the 17 linkage groups (LGs) covering 1,582.5 centimorgan (cM) with an average distance of 1.1 cM. The longest LG with a genetic distance of 111.9 cM was LG8 where 144 markers including 1 SSR were located, while the shortest LG with a genetic distance of 74.6 cM was LG10, where 120 markers including 1 SSR were loaded. Especially, both 1 SSR and 1 InDel were coalesced to LG6. Finally, 12 out of 17 SSR markers were successfully fixed on LG2, 5, 6, 8, 10, 13, 16, and 17 whereas 5 out of 13 InDel markers were fixed on LG4, 6, 7, and 14. The results of our study could be used as a basic frame map for comparative analysis of genomic structure between different research groups in pears.

P0692: Fruit Species

LD Analysis and Phylogeny Study in Citrus with High-Density SNP Array Data

Yoko Eck[1], Sergio Pietro Ferrante[2], Mikeal L. Roose[2], Marcos Moragues Canela[3], Claire Federici[1] and Timothy J. Close[4], (1)University of California Riverside, Riverside, CA, (2)University of California, Riverside, CA, (3)Univerisity of California at Riverside, Riverside, CA, (4)Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA

We performed linkage disequilibrium (LD) and phylogeny analyses with high-density SNP genotyping data generated by a recently developed SNP genotyping array for *Citrus*, Axiom™ Citrus 56AX (Affymetrix, Inc.) (58K autosomal and 500 Chloroplast SNPs). Recent technological advancements, such as array technology, not only significantly increased the amount of data available, but also provided increased throughput, in order to perform high-resolution genomic studies of an entire germplasm collection that would have been otherwise difficult to perform. To show the utility of Axiom Citrus 56AX, we performed LD and phylogeny analyses using 878 unique accessions from UCR Citrus Variety Collection (CVC). The values of LD in subgroups and population structure must be carefully examined to assess the appropriate marker density and to avoid spurious associations prior to performing GWAS. Studying LD and phylogeny in this germplasm will allow us further understand the relationships among these accessions and will be useful for further analyses in the future. We used most stringent PolyHighResolution (PHR) loci as classified by Axiom™ Analysis Suite in both analyses. The phylogenetic analyses with whole genome sequence data of 38 accessions (nuclear and chloroplast) used as a part of citrus variant discovery panel for the array design were also performed. The results obtained were then compared to those produced by array genotyping data to show that the array genotyping data is comparable to the sequence data. Our studies show the utility of a high density SNP array for LD analysis and constructing phylogenetic trees of *Citrus*.

P0693: Fruit Species

Genome-Wide Association Study Identifies Natural Alleles Associated with HLB Tolerance in Citrus

Qibin Yu[1], Amany Mira[2], Mattew Matthew[2], Honghong Deng[3], Zhibin WANG[4], Ming Huang[1], Dongliang Du[1], Yi Zhang[1], Zhanao Deng[5] and Fred G. Gmitter[1], (1)University of Florida, IFAS-CREC, Lake Alfred, FL, (2)University of Florida, Lake Alfred, FL, (3)Southwest University, College of Horticulture and Landscape Architecture, Chongqing, China, (4)UNIVERSITY OF FLORIDA, LAKE ALFRED, FL, (5)University of Florida, Wimauma, FL

Commercial citrus varieties have little resistance to Huanglongbing (HLB), a disease presumably caused by *Candidatus* Liberibacter asiaticus (*C*Las). As HLB has become widespread in Florida over ten years, we have very carefully followed the extensive collection of existing hybrid families, somaclonal variants, induced mutants, and germplasm accessions within the UF-CREC citrus breeding program. Identification of natural alleles associated with HLB tolerance is critical to the development of *C*Las tolerant citrus cultivars to combat the disease. A mandarin breeding population (n=192) was genotyped using an Axiom citrus56 SNP Array. The linkage disequilibrium and population structure of mandarin breeding population were investigated. Through a genome-wide association study (GWAS), we report identification of natural alleles in several candidate genes, such as HLB susceptible protein and HLB early signaling defense response, associated with tolerance. Overall, our studies indicated that HLB tolerance is a complex trait, and potentially could be exploited by using advanced breeding tools such as GWAS and genomic selection.

P0694: Fruit Species

Identification, Characterization and Expression Analysis of Lineage-Specific Genes within Sweet Orange (*Citrus sinensis*)

Yuantao Xu, Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China

Citrus is an important fruit crop worldwide and have immense economic value. With the availability of rapidly increasing number of genome and transcriptome sequences, lineage-specific genes (LSGs) can be identified and characterized. We compared the sweet orange (one major citrus cultivated species) genome sequences with 41 plant genomes (released in Phytozome v10.1 excluding *Citrus clementina* and *C. sinensis* (sweet orange) and 273 trancriptomes (PlantGDB-assembled Unique Transcripts) and identified two set of LSGs: 296 citrus-specific genes (CSGs, genes for which we could find at least one homolog in citrus, but no homologs anywhere else) and 1,039 orphan genes (genes for which we could not find homologs in any other species) specific to sweet orange. With the two sets of genes, gene structure and gene expression pattern were investigated. On average, both the CSGs and orphan genes have fewer exons, shorter gene length and higher GC content when compared with those evolutionarily conserved genes (ECs). Expression profiling indicated that most of the LSGs expressed in various tissues of sweet orange and some of them exhibited distinct temporal and spatial expression patterns. Besides, part of the CSGs and orphan genes expressed responsive to abiotic stresses. This study identified and characterized two sets of citrus-specific genes, dissected their sequence features and expression patterns, and provided a firm ground of citrus specific gene resources and useful clues for future dissection of the functions of citrus-specific genes to understand the specific biology in citrus

P0695: Fruit Species

Identification of CHS Superfamily Genes in Mandarin-Blood Orange Hybrid Seedlings after Methyl Jasmonate Treatment Reveals New Relationship between Gene Expression and Flavonoid Biosynthesis

Zhibin Wang[1], Qibin Yu[1], Wanxia Shen[2], Choaa Amine A Ei-Mohtar[1], Xiaochun Zhao[2] and Fred G. Gmitter[1], (1)University of Florida, IFAS-CREC, Lake Alfred, FL, (2)Southwest University, Chongqing, China
Citrus flavonoids are important because of their various biological and pharmacological activities. The chalcone synthase (CHS) gene superfamily consist of type III polyketide synthases (PKSs) and catalyze the first committed step in flavonoid biosynthesis. Although many studies related to *CHS* gene and its function in the flavonoid pathway have taken place, the *CHS* superfamily and the accurate relationship between gene expression and flavonoids accumulation, especially the functional gene identification, are still unclear in citrus.
In the current study, 77 potential *CHS* or *CHS* related genes or CDS sequences were screened to identify functional genes in seedlings of a mandarin-blood orange hybrid. Ten candidate members were used for qPCR analysis and 3 members were studied in overexpression and RNAi transgenic experiments. A new novel gene (MF776052) was identified and cloned; it has four CHS-specific conserved motifs and a CHS-family signature sequence GFGPG.
The results showed that the *CHS* gene superfamily should contain 5 or more members: CICLE_v10015535mg, CICLE_v10001405mg, CICLE_v10028604mg, CICLE_v10005133mg and >XM_006487413.1. At least three functional *CHS* gene members exist in this hybrid, with distinct contributions to flavonoid biosynthesis. The functional genes may be used in pathway engineering to increase flavonoid production for human consumers.

P0696: Fruit Species
Construction of High Density Genetic Maps and Detection of QTLs Associated with Titer of '*Candidatus* Liberibacter asiaticus' in Citrus
Ming Huang[1], Mikeal L. Roose[2], Qibin Yu[1], Dongliang Du[1], Yi Zhang[1], Zhanao Deng[3], Michael Irey[4], Ed Stover[5] and Fred G. Gmitter[1], (1)University of Florida, IFAS-CREC, Lake Alfred, FL, (2)University of California, Riverside, CA, (3)University of Florida, Wimauma, FL, (4)U.S. Sugar Corporation, USDA, Clewiston, FL, (5)U.S. Horticultural Research Laboratory, USDA-ARS, Fort Pierce, FL
No true resistance to Huanglongbing (HLB), a disease associated with infection of citrus by *Candidatus* Liberibacter asiaticus (*C*Las), is found within commercial cultivars, though trifoliata (*Poncirus trifoliata*) has been described as resistant or tolerant. Utilizing the Genotyping-by-Sequencing approach, a SNP-based genetic linkage map was constructed separately for trifoliata and sweet orange (*Citrus sinensis*) using an intergeneric F1 population of 170 individuals. For trifoliata, 647 high-quality SNP markers with unique loci were mapped into nine linkage groups, spanning a total genetic length of 1030.8 cM with average inter-loci distance of 1.59 cM. For sweet orange, 754 high-quality SNP markers with unique loci also were mapped into nine linkage groups, spanning a total genetic length of 760.2 cM with average inter-loci distance of 1.01 cM. Genetic maps of both species showed high linear correlation with the Clementine mandarin genome. After being exposed to intense HLB pressure for two years in a field trial, a collection of 113 individuals with 8 clonal replicates per individual, including 688 F1 progenies, 48 trifoliata and 16 sweet orange, were phenotyped for *C*Las titer in leaves by qPCR from October 2013 to October 2016 three times per year. The mean Ct value associated with *C*Las titer through all the time points was above 37 in all trifoliata varieties, but under 26 in all sweet orange varieties, while it ranged from 26 to 37 in F1 progenies. The Ct values and the percentage of healthy trees were associated with the genetic maps to detect QTLs respectively for trifoliata and sweet orange at each of time point and in each year. Combining the QTLs of both traits at different times, generally six significant QTLs were detected in trifoliata, mostly mapped on LG3, LG7 and LG8, explaining 16-22% of variance; while eight QTLs were detected in sweet orange, mostly mapped on LG3, LG5 and LG9, explaining 17-33% of variance. The final identified QTLs could be good targets for breeding to support long-term control of this devastating disease.

P0697: Fruit Species
High-Density Linkage Maps for *Citrus sunki* and *Poncirus trifoliata* using DArTseq Markers
Maiara Curtolo[1], Tatiany Soratto[1], Rodrigo Gazaffi[2], **Marco A. Takita**[1], Mariangela Cristofani-Yaly[1] and Marcos A. Machado[1], (1)Sylvio Moreira Citrus Center, Agronomy Institute, Cordeiropolis, Brazil, (2)Federal University of Sao Carlos, Araras, Brazil
The construction of a high-resolution genetic map of citrus would be of great value to breeders and to relate genomic regions with characteristics of agronomic interest. We constructed a novel high-resolution map of citrus using a population derived from a controlled cross between *Citrus sunki* (female parent) and *Poncirus trifoliata* (male parent). The genetic linkage maps were constructed using DArTseq markers and a pseudotestcross strategy; only markers showing the expected segregation were considered. To investigate synteny, all markers from the two linkage maps were aligned with the genome of *Citrus sinensis.* The *C. sunki* map has a total of 2,778 molecular markers and a size of 2,446.6 cM, distributed across ten linkage groups. The map of *P. trifoliata* was built with 3,084 markers distributed in a total of nine linkage groups, with a total size of 2,411.55 cM. These maps are the most saturated linkage maps available for *C. sunki* and *P. trifoliata* and have high genomic coverage. Both maps reported here are closely related to the genome of *C. sinensis*.
Financial support: CNPq (465440/2014-2) and FAPESP (2014/50880-0)

P0698: Fruit Species
Re-Sequencing and RNA-Sequencing of 'LB8-9' Sugar Belle® Mandarin and its Parents
Dongliang Du[1], Honghong Deng[2], Chunxian Chen[3], Qibin Yu[1], Ming Huang[1], Yi Zhang[1], Nidhi Rawat[4], Zhanao Deng[5], Fred G. Gmitter[1], Saroj Parajuli[4] and Xiaobao Ying[4], (1)University of Florida, IFAS-CREC, Lake Alfred, FL, (2)Southwest University, College of Horticulture and Landscape Architecture, Chongqing, China, (3)USDA, ARS, SEFTNRL, Byron, GA, (4)University of Florida, IFAS-GCREC, Wimauma, FL, (5)University of Florida, Wimauma, FL
Huanglongbing (HLB) is considered the most destructive disease of citrus because it causes tree decline and reduced fruit yield and quality, and affects all cultivars. Since its spread in Florida, HLB has cut the citrus production by more than 70% in ten years. With the continuous spread of HLB in Florida and worldwide, there is an urgent need for HLB-tolerant cultivars. Field observations showed that the growth and yield of 'LB8-9' Sugar Belle® mandarin were not suppressed by HLB, while its parents ('Clementine' mandarin and 'Minneola' tangelo) were more sensitive to HLB. In this study, whole genome re-sequencing and RNA-sequencing analyses of 'LB8-9' and its parents were conducted to find the possible reasons behind its HLB tolerance. More than 800 genes were identified containing high-effect SNPs associated with 'LB8-9'

genotypes by SNPeff. RNA-seq results showed that 1,416 genes were differently expressed between 'LB8-9' and 'Clementine'. The most enriched GO biological terms included plant-type secondary cell wall biogenesis, response to bacterium, and transmembrane receptor protein tyrosine kinase signaling pathway.

P0699: Fruit Species

The Relationship between Huanglongbing Symptom Severity and the Transcriptional and Anatomical Changes within Lemon (*Citrus limon* (L.) Osbeck) Trees

Honghong Deng[1,2], Dongliang Du[2], Diann S Achor[2], Qibin Yu[2], Guolu Liang[1] and Fred G. Gmitter[2], (1)Southwest University, College of Horticulture and Landscape Architecture, Chongqing, China, (2)University of Florida, IFAS-CREC, Lake Alfred, FL

Huanglongbing (HLB) is an extremely destructive and lethal disease of citrus worldwide presumably caused by a phloem-limited bacteria, *Candidatus* Liberibacter asiaticus (*C*Las). A better understanding of HLB disease is required before improvements can be targeted. In this study, we compared the global gene expression and anatomical aberrations of symptomatic and asymptomatic leaf samples from *C*Las-infected lemon trees, with the objective of evaluating transcriptional and anatomical changes concomitant with HLB symptom development. RNA-Seq results identified 778 and 603 differentially expressed genes (DEGs) in the symptomatic versus asymptomatic leaves of lemon trees exhibiting minor or severe decline symptoms, respectively. Compared to asymptomatic leaves, symptomatic leaves showed significant downregulation of genes associated with enzyme inhibitor activity, carbon-oxygen lyase, and cell wall-related genes such as cellulose synthase; these categories of genes are consistent with anatomical observations of abnormal massive starch accumulation and phloem collapse with cell wall distortion and breakdown. UDP-glucosyltransfer and other transferases in asymptomatic leaves were up-regulated which was related to transmembrane transporting. Intriguingly, for the symptomatic leaves, only 18 DEGs were found between minor and severe decline lemon trees, between which similar results were found for asymptomatic leaves with only 4 DEGs. The small differences between lemon trees with minor or severe decline may imply the ability of lemon trees to produce new healthy flush annually.

P0700: Fruit Species

Identification of Female Component of Self-Incompatibility in *Citrus grandis* Osbeck

Mei Liang[1], Lijun Chai[1] and Xiuxin Deng[2], (1)Huazhong Agricultural University, wuhan, China, (2)Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China

Self-incompatiblity (SI) is a widespread system adopted by almost flowering plants, which strictly inhibites inbreeding and promotes outcrossing. In citrus, the vast majority pummelo varieties express such mechanism to reject its own pollen. Despite dozens of components involved in SI system were well documented, it remains unclear what, if at all, determine and control the SI reaction in pummelo. In this study, we present eight pummelo varieties with SI trait and indicate that a relative higher ribonuclease activity exists in their stylar proteins, compared to those from papavariceae. From the differential expression profiles between style and anther from 'Shatian' pummelo, we annotate two ribonuclease T2 genes, named *CgRNS6* and *CgRNS9*, which both possess the certain genetic features of female components in SI, such as intron pattern, isoelectric point, conserved domain, and style-specific expression. Among the F1 population of 'Shatian' pummelo and Suan pummelo, whose genotypes are S6S9 and S9Sn ('n' represents unknown genotype), the segregation ratio of CgRNS6 and CgRNS9 is 1:1, and the pollens with S9 are all inhibited because there are no S6S9 or S6S9 plants to be identified. Finally, the two recombinant proteins are isolated from *E. coli*, and such proteins significantly inhibited the growth of self-pollen tubes in the in vitro culture system. In conclusion, we confirmed that *Citrus* also shared the S-RNase-based mechanism, which represents a major advance in our understanding on the SI system of *Citrus*.

P0701: Fruit Species

Genomic Analyses Provides a Insight to Fruit Acidity Change during Mandarin Domestication

Lun Wang, Huazhong agricultrual University, Wuhan, China

The domestication of fruit crops was more recent than the domestication of annual crops. The prominent traits that were selected during the domestication of fruit trees include fruit flavor, sweetness/acidity, fruit morphology, and nutritional qualities.The wild mandarins contained remarkably high levels of citric acid relative to cultivated forms. We identified the most pure (i.e., least admixture) mandarin, and found evidence that two independent domestication events occurred to yield one group with changes in genes associated with primary metabolism and another group with changes in genes associated with fruit acidity and color.

P0702: Fruit Species

Genomic Analyses of Primitive, Wild, and Cultivated Citrus Provide Insights into Citrus Polyembryony

Xia Wang[1], Yuantao Xu[1], Siqi Zhang[1], Li Cao[2], Xiuxin Deng[1] and Qiang Xu[1], (1)Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China, (2)Citrus Research Institute, Southwest University, Chongqing, China

Our research group has *de novo* assembled five citrus genomes representing primitive, wild and cultivated citrus. Among them, the most completed citrus genome was generated from haploid pummelo by single-molecule sequencing on PacBio platform with the contig N50 reaching 2.2 Mb. We further sequenced 100 accessions of primitive, wild, and cultivated citrus, which were collected by our group. Comparative population analyses displayed the highest genetic diversity among primitive citrus accessions, and that genomic regions harboring reproduction- and energy-related genes are probably under selection in cultivated citrus.

In citrus, nucellar polyembryony is a unique apomixis phenomenon, in which the embryos develop from somatic nucellar cells. Polyembryony is widely employed in citrus nurseries and propagation programs to generate large numbers of uniform rootstocks from seeds and to permanently fix valuable traits and hybrid vigor, while it has also caused problems for breeding that required sexual crosses, which make it of great value to study the citrus polyembryony. Because of long juvenile phase of woody citrus plants and citrus polyembryony, conventional genetic mapping methods by constructing hybrid population after several generations are not applicable for citrus. In our study, a strategy combing bulk segregant analysis with local gene-based association analysis was designed for genetic mapping in citrus. The final genetic locus

responsible for citrus polyembryony was narrowed to an 80-kb region containing 11 genes. One of these, *CitRWP*, shows the highest association with the polyembryony phenotype and is expressed at higher levels in ovules of polyembryonic cultivars. We found a miniature inverted-repeat transposable element insertion in the promoter region of *CitRWP* that cosegregated with polyembryony. This study provides new insights into citrus polyembryony and provides practicable references for genetic mapping of agriculturally important genes in citrus.

P0703: Fruit Species

Identification of Candidate Genes for Chlorophyll Mutant in Pummelo (*Citrus maxima*)

Huiwen Yu, Key Laboratory of Horticultural Plant Biology, Huazhong Agricultural University, Wuhan, China and Xia Wang, Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, Wuhan, China

Bud sport is widely used in citrus breeding, however, it is difficult to distinguish from normal ones. Furthermore, the natural variation on leaf color is rare in citrus. A chlorophyll bud sport of 'Guanxi' pummelo (*Citrus maxima*) which showed etiolated leaves at young stage was used to construct segregating population in the study. The extreme phenotype of F1 generation with normal green leaves and etiolated leaves were selected to construct pools. Bulked segregant analysis (BSA) and indel markers identified a 0.3Mb region on chromosome 6. Gene expression analysis of bulk RNA further identified four candidate genes that were significantly differential expression both in parental lines and F1 bulks. Gene sequence analysis of these four genes indicated there were two indel differences in *Chl570* between the green and etiolated generations. This study provides a new germplasm of pummelo for citrus breeding and the candidate gene may has important function in chloroplast biological processes.

P0704: Fruit Species

Functional Genomics of Citrus Pathogens Focusing on Effectors Biology

Marcos Antonio Machado[1], Ronaldo JD Dalio Dalio[2], Eduardo Henrique Goulin[2], Carolina Munari Rodrigues[2], Paulo José Camargo dos Santos[2] and Marco Aurélio Takita[3], (1)Centro de Citricultura/IAC, Cordeiropolis, Brazil, (2)Centro de Citricultura/IAC, Cordeirópolis - SP, Brazil, (3)Centro de Citricultura/IAC, Cordeiropolis - SP, Brazil

Diseases caused by fungi and oomycetes causes serious problems in the production of citrus worldwide. Pathogens that affect mainly the rootstock, such as root rot and gummosis caused by Phytophthora spp, or the production or quality of the fruits as post-bloom fruit drop (PFD) caused by Colletotrichum abscissum, black spot (Phyllosticta citricarpa), and brown spot (Alternaria alternata) represent a significant increase in production costs. The molecular mechanisms of interactions of these pathogens with citrus are still unknown. With the sequencing of their genomes, an approach to understanding these interactions is the search for candidate genes to act as effectors, i.e., to produce proteins that are secreted out of the cell and capable of altering host structure and physiology. Since they are fundamental virulence factors of the pathogens, management strategies for disease control, specially to be used as probes to screen for genome and transcriptome data from these pathogens, a bioinformatic pipeline was developed for selection of candidate effectors genes based on criterious, such as presence of signal peptides for secretion, presence or absent of transmembrane alfa-helices, secondary and tertiary structure. All candidate genes are re-sequenced and some of them have been functionally evaluated in model plants for induction of hypersensitivity responde (HR). The most promising are being used in the transformation of Arabidopsis to confirm or not their ability to alter host physiology. Financial support: INCT Citros (Fapesp 2014/50880-0 and CNPq 465440/2014-2).

P0705: Fruit Species

Liposome Delivery System of Antimicrobial Peptides against Huanglongbing (HLB) Citrus Disease

Jeanette Velasquez[1], Supratim Basu[1], Roel C. Rabara[1], Loan Huynh[1], Hau Nguyen[1], Goutam Gupta[1], Shujian Zhang[1] and U. S. Horticultural Research Laboratory, (1)New Mexico Consortium, Los Alamos, NM

Previously, we have shown that engineering of a novel immunity facilitates rapid clearance of pathogens from the site of colonization in plants. This is achieved by generation of a disease-resistant transgenic plant that expresses a therapeutic protein capable of recognition and lysis of the invading pathogen. Here, we focus on plant amphipathic helical peptides that can only kill bacteria with poor activity and low specificity and are often susceptible to bacterial resistance. Our target system is *Candidatus Liberibacter*, which causes Huanglongbing, the most devastating disease of citrus. Toward this end, based upon endogenous peptides we have designed and synthesized eleven amino acid long membrane-targeting helical amphipathic peptides with antimicrobial activity against Gram negative bacteria. Since Liberibacter is not culturable, we have tested their activity on several gram-negative bacteria surrogates such as *Escherichia coli*, *Salmonella typhimurium*, *Agrobacterium tumefaciens*, and *Sinorhizobium meliottii*. The last two are Liberibacter surrogates. Our minimum inhibitory concentration (MIC) experimental results indicated that P11-1 peptide is the most active one on *E. coli, A. tumefaciens* and *S. typhimurium*. By an ex planta assay, we have shown that P11 and its analogs are capable of reducing the Liberibacter load in infected citrus leaves. In fact, P11 peptides are more active on Liberibacter than currently used streptomycin. In addition, we also examined how bacteria (such as *E. coli*) develop resistance against P11 thereby making it ineffective. This prompted us to design the 2nd generation helical amphipathic peptide P26, in which two P11 peptides are connected by a 4 amino acid beta turn. We have shown that P26 is not only more active on *E. coli* but also not susceptible to bacterial resistance. Ex planta assay also shows that P26 possesses higher anti-Liberibacter activity than P11 or streptomycin. Currently, we are testing the efficacy of P11 and P26 by in planta assays using two delivery systems. One delivery system utilizes the spray of naked P11 and P26 in detergents whereas the other uses virus-like liposome vesicles that encapsulate P11 and P26. Attachment of virus coat protein on liposome vesicles should ensure transport in the plant, cell to cell transfer, and controlled release of P11 and P26. Successful completion of these in planta studies would open the possibility of using P11 and P26 for HLB treatment.

P0706: Fruit Species

Molecular Pathways Controlling the "Shape" of Persimmon Fruits

Haruka Maeda[1], Takashi Akagi[1,2] and Ryutaro Tao[1], (1)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (2)JST-PRESTO, Saitama, Japan

Plant organ shapes reflect not only their developmental evolutions but also a commercial quality in crops. There have been many reports about molecular pathways involving leaf shapes in *Arabidopsis* and tomato, mainly from a viewpoint of evolutionary developmental biology. On the other hand, little is known about the pathways controlling fruit shapes, despite their wide diversity in many crops. Persimmon (*Diospyros kaki*) cultivars show a wide diversity in fruit shape. We have developed a method for quantification of persimmon fruit shapes by combining elliptical Fourier descriptors and principal component analysis. Here, we aimed to elucidate the molecular mechanism controlling the diversity of the persimmon fruit shapes, by assembling transcriptomic data and the quantitative scores of the shapes. Chronological characterization of quantitative shape scores in over 150 persimmon cultivars unveiled the fundamental stage contributing the diversity of the fruit shapes. Transcriptomic data using representative 50 cultivars during the stage important for fruit shape determination were assessed to detect the association with the standardized variations of the main principal components (PCs) of the shape scores, by Pearson-momental correlation analysis. The results suggested that some candidate genes potentially related to cell wall and/or cell differentiation, such as *EXPA* and *KAN1*, can be partially involved in regulation of the fruit shape diversity in the persimmon cultivars.

P0707: Fruit Species

RNA-Seq Analysis of the Mango (*Mangifera indica* L) Fruit Epidermis: Elucidating the Molecular Mechanism of Cuticle Biosynthesis

Martín-Ernesto Tiznado-Hernández, Coordinación de Tecnología de Alimentos de Origen Vegetal, Hermosillo, Mexico and Tafolla-Arellano, Julio César, Zheng, Yi; Sun, Honghe; Jiao, Chen; Ruiz-May, Eliel; Hernández-Oñate, Miguel A.; González-León, Alberto; Báez-Sañudo, Reginaldo; Fei, Zhangjun; Domozych, David; K.C. Rose, Jocelyn

México is one of the greatest producer and exporter of mango fruit (*Mangifera indica* L.), which are highly perishable and have a limited shelf life and global distribution, mainly due to postharvest desiccation and senescence. Several treatments had been developed to prolong the postharvest shelf life, however, it seems that those treatments had reached the maximum limit to increase the time of the postharvest shelf life of horticultural products. The cuticle is the outermost layer of the plant cell and it is composed of waxes and cutin, including long chain fatty acids, phenolic compounds as well cell wall polysaccharides. The cuticle is an interface between the fruit and the environment, the primary role is to control the water transpiration, also it is involved in plant-insect interaction and pathogen attack. In this context, evidences from studies with a tomato mutant, strongly indicates that the cuticle plays an important role in the postharvest shelf life of fruits. However, studies of these phenomena in mango fruit are limited by the lack of genome-scale data. In order to gain insight into the mango cuticle biogenesis and identify putative cuticle-associated genes, we analyzed the transcriptomes of peels from ripe and overripe mango fruit using RNA-Seq. Approximately 400 million of reads were generated and *de novo* assembled into 107,744 unigenes, with a mean length of 1,717 bp. Out of these, 7740 were annotated in 461 metabolic pathways, 5342 transcription factors and 3662 protein kinases. With the goal to distribute our transcriptome sequences to the research community and allow researchers to mine the mango transcriptome dataset, we developed an online database called Mango RNA-seq Database (http://bioinfo.bti.cornell.edu/cgi-bin/mango/index.cgi). RNA-Seq analysis suggested that the pathway leading to biosynthesis of the cuticle component, cutin, is up-regulated during overripening. This data was supported by analysis of the expression of several putative cuticle-associated genes and by gravimetric and microscopic studies of cuticle deposition, revealing a complex continuous pattern of cuticle deposition during fruit development and involving substantial accumulation during ripening/overripening.

P0708: Fruit Species

Adventitious Rooting in Avocado - Unravelling the Bottleneck to Propagation

Madeleine Gleeson, **Alice Hayward**, Jayeni Hiti-Bandaralage, Stephen Fletcher and Neena Mitter, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia

Avocado (*Persea americana* Mill.) is a highly nutritious fruit of immense popularity. As orchards expand to meet consumer-driven increases in demand, the rapid propagation of elite cultivars is of the utmost importance. The clonal propagation of avocado rootstocks however, is a significant bottleneck to the industry due to the immense recalcitrance to adventitious root (AR) generation. This recalcitrance hinders both existing nursery-based propagation, in addition to the development of emerging tissue culture pipelines. We have observed that in tissue culture, shoots of some cultivars generate AR much more readily than those of others, including some of the more industrially relevant cultivars. To examine why one cultivar was AR responsive to a treatment in which the other cultivar was not, an RNAseq with qPCR validation was completed. Plant material was sequenced from tissue cultured shoots of an easy-to-root (AR+) and a difficult-to-root (AR-) cultivar 72 hours after treatment with either mock or root-inducing substrates. A selection of genes significantly differentially regulated between treatments and across cultivars were selected for validation by qPCR. These genes were also profiled across additional treatments and time points (0 hours and 24 hours) for each cultivar, to further dissect their relationship to rooting phenotype. The molecular profiles of these genes, many of which have not previously been implicated in AR, provide invaluable insight into avocado AR regulation and will guide improvements to rooting protocols into the future.

P0709: Fruit Species

Identification of Avocado Cultivars using Genomic and Morphometric Tools

Alicia Talavera[1], Eduardo Palma-Martín[2], Aboozar Soorni[3], Therese Bruwer[4], Zelda van Rooyen[4], Aureliano Bombarely[3], Antonio J. Matas[1,2] and J. Inaki Hormaza[1], (1)IHSM "La Mayora" - UMA - CSIC, Málaga, Spain, (2)Departamento de Biología Vegetal. Universidad de Málaga, Málaga, Spain, (3)Virginia Polytechnic Institute and State University, Department of Horticulture, Blacksburg, VA, (4)Westfalia Technological Services, Tzaneen, South Africa

Subtropical fruits, like avocado, are key crops for food security in a wide range of countries, with an increasing commercial importance worldwide. However, the lack of accurate molecular and phenotypic data is a critical bottleneck for appropriate breeding and selection of new cultivars in these crops. Thanks to next-generation sequencing approaches we now have the ability of developing a high number of molecular markers under reasonable costs. In addition, the utilization of new phenotyping techniques are efficient tools to overcome the difficulties of the limited previous genetic and phenotypic information available for these fruit crops.

In this work, we have generated a collection of SNP markers using 93 avocado previously phenotyped cultivars. The results provide a new set of tools for an efficient avocado variety characterization and optimization of avocado breeding programs to obtain new cultivars for the diversification of the avocado market.

P0711: Fruit Species

Papain-like Cysteine Proteases in *Carica papaya*: Lineage-Specific Gene Duplication and Expansion

Juan Liu[1,2], Anupma Sharma[2], Marie Jamille Niewiara[3], Ratnesh Singh[2], Ray Ming[3] and Qingyi Yu[2], (1)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology;Fujian Agriculture and Forestry University, Fuzhou, China, (2)Texas A&M AgriLife Research, Dallas, TX, (3)University of Illinois at Urbana-Champaign, Urbana, IL

Background. Papain-like cysteine proteases (PLCPs), a large group of cysteine proteases structurally related to papain, play important roles in plant development, senescence, and defense responses. Papain, the first cysteine protease whose structure was determined by X-ray crystallography, plays a crucial role in protecting papaya from herbivorous insects. Except the four major PLCPs purified and characterized in papaya latex, the rest of the PLCPs in papaya genome are largely unknown.

Results. We identified thirty-three PLCP genes in papaya genome. Phylogenetic analysis clearly separated plant PLCP genes into nine subfamilies. PLCP genes are not equally distributed among the nine subfamilies and the number of PLCPs in each subfamily does not increase or decrease proportionally among the seven selected plant species. Papaya showed clear lineage-specific gene expansion in the subfamily III. Interestingly, all four major PLCPs purified from papaya latex, including papain, chymopapain, glycyl endopeptidase and caricain, were grouped into the lineage-specific expansion branch in the subfamily III. Mapping PLCP genes on chromosomes of five plant species revealed that lineage-specific expansions of PLCP genes were mostly derived from tandem duplications. We estimated divergence time of papaya PLCP genes of subfamily III. The major duplication events leading to lineage-specific expansion of papaya PLCP genes in subfamily III were estimated at 48 MYA, 34 MYA, and 16 MYA. The gene expression patterns of the papaya PLCP genes in different tissues were assessed by transcriptome sequencing and qRT-PCR. Most of the papaya PLCP genes of subfamily III expressed at high levels in leaf and green fruit tissues.

Conclusions. Tandem duplications played the dominant role in affecting copy number of PLCPs in plants. Significant variations in size of the PLCP subfamilies among species may reflect genetic adaptation of plant species to different environments. The lineage-specific expansion of papaya PLCPs of subfamily III might have been promoted by the continuous reciprocal selective effects of herbivore attack and plant defense.

P0712: Fruit Species

Genome-Wide Comparative Analysis between PRSV Resistant Transgenic Papaya and its Progenitor Cultivar Reveals Structural Variations Induced by Biolistic Based Transformation

Jinping Fang, Fujian Normal University, Fuzhou, China

The safety of genetically transformed plants remains a subject of scrutiny. Data gathered on the structural variation in PRSV resistant transgenic papaya induced by miroprojectile-mediated transformation will provide a scientific-based means to rationally address such concerns. Using paired-end short-read next-generation sequencing (NGS), a total of more than 74 million reads for nontransgenic papaya Sunset were obtained and mapped onto transgenic papaya SunUp reference genome. In total, 310,364 polymorphisms (SNPs), 34,071 Small InDels (Inserts/deletions) and 1,200 structural variations (SVs) were found between Sunset and SunUp. Those variations have an uneven distribution across the 9 chromosomes of papaya. Regarding the effects of the mutations on gene function, >90% of the mutations were located in intergenic regions, while only 0.27% were predicted to be high-impact mutations. Gene ontology (GO) enrichment analysis revealed that ATP-related categories were highly enriched among these high-impact genes. The SNP mutation rate was about 8.4×10^{-4} per site, highly comparable with the rate induced by spontaneous mutation over numerous generations. The transition-to-transversion ratio was 1.439 and the predominant mutations were C/G to T/A transitions. We inferred that spontaneous mutation was the leading cause of SNPs in transgenic papaya SunUp. We also studied the integration of organelle DNA into the nuclear DNA of SunUp and Sunset. A total of 3,430 nuclear plastid DNA (NUPT) and 2,764 nuclear mitochondrial DNA (NUMT) junction sites had been found in SunUp, which is proportionally higher than the predicted total NUPT and NUMT junction sites in Sunset. Among all nuclear organelle DNA (norgDNA) junction sites, 95% junction sites were shared by SunUp and Sunset, suggesting norgDNA are abundant in papaya nuclear genomes and highly conserved after transgene insertion. NUPT/NUMT junction sites exclusively in SunUp were significantly different from those in Sunset. The average identity between SunUp specific norgDNA and corresponding organelle genomes was higher than that of norgDNA shared by SunUp and Sunset. BLAST between the six SunUp organelle-like borders of transgenic insertions and corresponding organelle genomes brought us to percent identities of 98.18~100%, which were higher than the identities between five NUPT borders and Sunset norgDNA. All paired-end spans of mapped Sunset reads were shorter than SunUp transformation plasmid derived inserts. Those hints suggested that many new DNA transferred from organelles to the nuclear genome during bombardment, including these six organelle-like borders. Taken together, the present study of comparative whole-genome analyses between SunUp and Sunset using NGS provides a reliable estimate of genome-wide discrepancy. Development of SNP/InDel markers that occurred in high-impact genes could facilitate marker-assisted PRSV disease resistance breeding in papaya. The newly integrated norgDNA induced by particle bombardment revealed the mechanisms underlying the process of foreign gene transformation.

Keywords: *Carica papaya* L.; Whole-genome resequencing; Structural variation; Nuclear plastid DNA (NUPT); Nuclear mitochondria DNA (NUMT)

P0713: Fruit Species

***Cucurbita* Genome Sequences Provide Insights into Polyploid Genome Evolution and Heterosis in Interspecific Hybrid**

Shan Wu[1], Honghe Sun[2], Guoyu Zhang[2], Chen Jiao[1], Shaogui Guo[2], Yi Ren[2], Jie Zhang[2], Haiying Zhang[2], Guoyi Gong[2], Zhangcai Jia[2], Fan Zhang[2], Jiaxing Tian[2], William Lucas[3], Jeffrey J. Doyle[4], Haizhen Li[2], Zhangjun Fei[1,5] and Yong Xu[2], (1)Boyce Thompson Institute, Cornell University, Ithaca, NY, (2)National Engineering Research Center for Vegetables, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China, (3)Deptartment of Plant Biology, Davis, CA, (4)Plant Biology Department, Cornell University, Ithaca, NY, (5)USDA-ARS Robert W. Holley Center for Agriculture and Health, Ithaca, NY

The *Cucurbita* genus in the Cucurbitaceae family contains several economically important species and interspecific hybrids between *C. maxima* and *C. moschata* are widely used as rootstocks for other cucurbit crops. We *de novo* assembled the genomes of *C. maxima* and *C. moschata*, which provided evidence supporting an allotetraploidization event in the *Cucurbita* lineage. We partitioned the genome into two homoeologous subgenomes based on different genetic distances to the species in the Benincaseae tribe, including melon, cucumber and watermelon. We estimate that the two diploid progenitors of *Cucurbita* successively diverged from Benincaseae around 31 and 26 million years ago (Mya), and the allotetraploidization happened earlier than 3 Mya, when *C. maxima* and *C. moschata* diverged. The subgenomes have largely maintained the chromosome structures of their diploid progenitors. During evolution, the two subgenomes have retained similar numbers of genes, and neither subgenome is globally dominant in gene expression. Such long-term karyotype stability and unbiased fractionation has not been commonly observed in other allopolyploid plants. These two high-quality genome sequences allowed us to detect transgressive gene expression changes in the *C. maxima* × *C. moschata* interspecific F_1 hybrid 'Shintosa' correlated with heterosis in important agronomic traits such as carotenoid content in fruits.

P0714: Fruit Species

QTL Mapping of Flowering Time, Fruit Size and Number in Populations Involving Andromonoecious True Lemon Cucumber

Yupeng PAN, Department of Horticulture, University of Wisconsin-Madison, MADISON, WI, Changlong Wen, Beijing Academy of Agricultural and Forestry Sciences, Beijing, China, Yuhui Wang, Horticulture Department, University of Wisconsin, Madison, WI, Xiaomao Cheng, Department of Horticulture, University of Wisconsin-Madison, Madison, WI and Yiqun Weng, USDA-ARS, Vegetable Crops Research Unit, Madison, WI

Andromonoecious sex expression in cucumber is controlled by the *m* locus, which encodes the 1-aminocyclopropane-1 –carboxylic acid synthase (ACS) in the ethylene biosynthesis pathway. This gene seems to have pleotropic effects on fruit size and number, but the genetic basis is unknown. The True Lemon cucumber is an andromonoecious heirloom bearing lemon-sized and nearly round fruits. A recombinant inbred population (RIL) was developed from the cross between True Lemon and the gynoecious cucumber inbred line WI2757, which was used in QTL mapping for flowering time (FT), fruit number (FN), mature fruit length (MFL) and diameter (MFD). Phenotypic data for the four traits were collected from 139 RILs in field trials across multiple environments. Two linkage maps were developed using 129 microsatellite markers as well as 1845 genome-by-sequencing (GBS)-based SLAF markers. QTL analysis identified 15 QTL for the four traits including 1 for FT (*ft1.1*), 3 for MFL (*mfl1.1*, *mfl1.2*, and *mfl6.1*), 5 for MFD (*mfd1.1*, *mfd1.2*, *mfd4.1*, *mfd5.1*, and *mfd6.1*), and six for FN (*fn1.1*, *fn3.1*, *fn4.1*, *fn5.1*, *fn6.1*, and *fn7.1*). Among these QTL, *ft1.1*, *mfl1.2*, *mfd1.2*, and *fn1.1* were all major-effect QTL for FT, MFL, MFD and FN, respectively, which were consistently co-localized at the *m* gene region on cucumber Chromosome 1 suggesting possible pleiotropic effects of the *m* locus on expression of these traits. The effect of andromonoecy on fruit size and fruit number in this population will be discussed.

P0715: Fruit Species

Analysis on the Molecular Mechanism of Photoperiod Regulating Flowering in Xishuangbanna Cucumber (*Cucumis sativus* L.*var. xishuangbannesis* Qi et Yuan)

Zhen Tian, Nanjing Agricultural University, Nanjing, China

Cucumber is an important vegetable in our daily life, and it belongs to day-neutral plant. Among them, Xishuangbanna cucumber is a strict short day plant, which needs certain short-day treatment before flowering in temperate region. However, the characteristics and molecular mechanisms of photoperiod regulating flowering in Xishuangbanna cucumber are still not clear. Here, RNA-Seq technology was used to find the molecular basis of genes in Xishuangbanna cucumber under different photoperiod treatments. Six sequencing libraries were prepared using the second leaves of plants exposed to long/short-day photoperiod treatments followed with various development stages. The expression patterns of ten differentially transcribed genes was validated using qRT-PCR. We obtained 1019 million clean reads, and 31,213 differential expression genes, among which 18,774 were significantly up-regulated and 12,439 were down-regulated. The expression levels of a great number of genes involved in photosynthesis, plant circadian rhythm and plant hormone signal transduction. These findings will facilitate for understanding the molecular mechanisms of floral transition process in Xishuangbanna cucumber.

P0716: Fruit Species

QTL Mapping for Downy Mildew Resistance from a *Cucumis hystrix* Introgression Line of Cucumber (*C.sativus* L.)

Kaijing Zhang, Nanjing Agricultural University, College of Horticulture, Nanjing, China

Downy mildew (DM) caused by *Pseudoperonospora cubensi* is the most important foliar disease of cucumber in many areas of the world. To date, the underlying molecular genetic mechanism of DM resistance in cucumber is still poorly understood. Recently, QTL-seq can be a rapid and efficient strategy for QTLs identification that takes advantage of bulked-segregant analysis (BSA) and next-generation sequencing (NGS). The present study employed the QTL-seq strategy to identify the major genomic region harbouring DM resistance QTL using a set of 155 F_6 recombinant inbred lines (RILs) population and confirmed the results by traditional QTL mapping analysis. Based on the phenotypic data of 155 RILs in 2016, QTL-seq identified a major genomic region (22.32-25.20 Mb) on Chr5 displaying an average Δ (SNP-index) = 1 for DM resistance. As for traditional QTL analysis for DM resistance in 155 RILs, we identified two QTLs (*dm5.1* and *dm5.2* with positive additive effect) in 2015 and six QTLs (*dm1.1*, *dm1.2* and *dm1.3* with negative additive effect, *dm5.1*, *dm5.2* and *dm6.1* with positive effect) in 2016. Among them, *dm5.1* and *dm5.2* were detected in two seasons repeatedly and the locus *dm5.2* was a stable major-effect QTL for DMR in two experiments. The QTL *dm5.2* interval (22.65-24.94 Mb) corresponded to the genomic region on chromosome 5 identified by QTL-seq covering 22.32-25.20 Mb. In the genomic region 22.32-25.20 Mb identified by QTL-seq, we found 5 nonsynonymous SNPs with Δ (SNP-index) = 1 in five candidate genes and two SNPs with Δ (SNP-index) = 1 in UTR3 of two genes. These seven genes will conduct the further analysis and the seven SNPs can be developed as the polymorphic markers for fine mapping of *dm5.2*.

P0717: Fruit Species

Identification of Interactions between S-Adenosylmethionine Synthetase with Two Protein Kinases in Cucumber

Meiwen He, College of Horticulture, Nanjing Agricultural University, Nanjing, China
S-adenosylmethionine synthetase (SAMs) is the critical enzyme that catalyzes *S*-adenosylmethionine (SAM) synthesis from methionine and ATP. SAM acts as a precursor to polyamines, ethylene and lignin and provides the methyl group for protein and DNA then contributes to plant development and stress response. Despite increasing knowledge of SAMs regulation at transcriptional levels, little is known about the post-translational regulation in plant cells. Based on its structure as a Ser/Thr site rich protein, it may interact with several protein kinases containing Ser/Thr kinase domain. In the present study, interactions of a F-box protein containing a kelch repeat motif (FBK) , and a calcium-development protein kinase (CDPK6) with *S*-adenosylmethionine synthetase were identified using a yeast two-hybrid system and bimolecular fluorescence complementary (BiFC) assays. We have found that *CsSAMs* gene expression is up-regulated by various stimuli. Interestingly, the changes of these two protein-coding genes were similar to the pattern of *CsSAMs*. Therefore, our results suggest that CsSAMs is involved in a variety of stress response pathways by interacting with several protein kinases. The data trigger a hypothesis that SAMs may function as a node of crosstalk among hormones and abiotic stresses.

P0718: Fruit Species

QTL Analysis of Variation in Cuticle and Epicuticular Wax Deposition in Cucumber Fruit

Stephanie Rett, Michigan State University, East Lansing, MI and **Rebecca Grumet**, Department of Horticulture, Michigan State University, East Lansing, MI
Cucumber (*Cucumis sativus*) fruit are covered by a layer of cuticle and epicuticular waxes that function to restrict water loss and prevent infection by pathogens. The cuticle is formed from a complex matrix of mainly cutin and waxes, whose precursors are synthesized and secreted by epidermal cells. Deposition of the cuticle and epicuticular waxes can vary among species, cultivars and developmental age. This study evaluates the natural variation of cuticle and epicuticular wax deposition in two cultivars of cucumber, for which draft genomes and a single nucleotide polymorphism (SNP) array are available: Chinese Long '9930'(CL), a Chinese fresh market cucumber, and 'Gy14,' a pickling type. Fruit were sampled from an S_7:S_8 Gy14 x CL recombinant inbred line (RIL) population (n=112) at 16 days post pollination (dpp), a time when most cuticle deposition has already taken place and strong differences for cuticle thickness are readily observed. Fresh tissue samples were cross-sectioned, stained for lipids with Sudan IV, and microscopically evaluated for: cuticle thickness, depth of cuticular intercalations, epidermal cell dimensions, and number and size of lipid bodies. A single gene model seems to largely explain the variation in cuticle thickness and intercalation depth, whereas, epidermal cell height appears to be quantitatively inherited. Intercalation depth co-segregated with epidermal cell height, but not cuticle thickness. Quantitative trait loci (QTL) analysis implicates factors on chromosome 1 contributing to several traits.

P0719: Fruit Species

An Overlooked Paleo-Tetraploidization in Cucurbitaceae and a Gold Standard to Decipher Complex Genomes

Jinpeng Wang[1], Yuxian Li[1] and Xiyin Wang[2], (1)North China University of Science and Technology, Tangshan, China, (2)UGA and NCST, Watkinsville, GA
Cucurbitaceae plants are of considerable biological and economic importance, and genomes of cucumber, watermelon, and melon have been sequenced. However, a comparative genomics exploration of their genome structures and evolution has not been available. Here, we aimed at performing a hierarchical inference of genomic homology resulted from recursive paleo-polyploidizations. Unexpectedly, we found that, shortly after a core-eudicot-common hexaploidy (ECH), a cucurbit-common tetraploidization (CCT) occurred, overlooked by previous reports. Moreover, we characterized gene loss (and retention) after these respective events, which were significantly unbalanced between inferred subgenomes, and between plants after their split. The inference of a dominant subgenome and a sensitive one suggested an allotetraploid nature of the CCT. Besides, we found divergent evolutionary rates among cucurbits, and after doing rate correction, we dated the CCT to be 90-102 million years ago, likely common to all Cucurbitaceae plants, showing its important role in the establishment of the plant family. Furthermore, we found that polyploidizations contributed to the expansion of key functional gene families. The present efforts laid a comparative genomics platform to support various researches in the Cucurbitaceae community and beyond.

P0720: Fruit Species

Two Ancestral Lineages of Melon Cultivars Revealed by the Analysis of *PolA1* Gene

Miku Sato[1], Arisa Tsuchiya[1], Katsunori Tanaka[2], Daisuke Ogawa[3], Kenji Kato[4] and Ikuo Nakamura[1], (1)Graduate School of Horticulture, Chiba University, Matsudo, Japan, (2)Hirosaki University, Hirosaki, Japan, (3)Institute for Horticultural Plant Breeding, Matsudo, Japan, (4)Okayama University, Okayama, Japan
Miku Sato[1], Arisa Tsuchiya[1], Katunori Tanaka[2], Daisuke Ogawa[3], Kenji Kato[4], Ikuo Nakamura[1] 1: Chiba University, 2: Hirosaki University, 3: Institute for Horticultural Plant Breeding, 4: Okayama University
Melon (*Cucumis melo* L.) is a complex species comprising several varietal groups, such as Cantalupensis, Reticulatus, Inodorus, Conomon, Makuwa, which are differentiated in fruits and seed traits. Thus, it is difficult to resolve the origin of melon by the analysis of phenotypic characters. We focused on a protein tag (Ptag) sequence in the C-terminal region of the largest subunit (POLA1) of RNA polymerase I. The Ptag sequence is highly conservative within a species but rapidly differentiated among species. Because *PolA1* gene is single copy in the melon genome and Ptag-coding sequence is scarcely recombined between allele in an interspecific hybrid, Ptag sequence of ancestral species is inherited to progeny without sequence change.
At first, we analyzed *PolA1* 19th intron sequence, within Ptag coding sequence, of melon cultivars and *Cucumis* species. The result showed that melon cultivars contained only two 19th intron sequences (H1 and H2 haplotypes), which were differentiated at 3 SNPs and 2 indels each other. The 19th intron sequences of other *Cucumis* species were distantly related to those of melon cultivars. In addition, Ptag sequences of melon cultivars were analyzed using RT-PCR. The results indicated that melon cultivars shared respectively corresponding Ptag sequences (H1 and H2 haplotypes), which were differentiated by two amino acid substitutions. Cultivars of Cantalupensis group contained H1 or H2 haplotype. Inodorus group had H2 haplotype while Reticulatus, Makuwa, and Conomon groups contained H1 haplotype.
These results suggested that melon cultivars were originated from two ancestral lineages in *Cucumis melo*.

P0721: Fruit Species

Construction of a High-Density Genome-Anchored Genetic Map for Melon (*Cucumis melo* L.) and Identification of *Fusarium oxysporum* f. sp. *melonis* Race 1 Resistance QTL

Sandra E. Branham[1], Amnon Levi[1], Melanie Katawczik[1], Zhangjun Fei[2] and **Patrick Wechter**[1], (1)USDA-ARS, Charleston, SC, (2)Boyce Thompson Institute, Cornell University, Ithaca, NY

The *Cucumis melo* inbred line MR-1, developed by the USDA, ARS possesses numerous resistance alleles for several major diseases of melon. These include powdery mildew, downy mildew, Alternaria leaf blight, and Fusarium wilt. A recombinant inbred population (RIL) was developed from an initial cross of MR-1 to a highly susceptible Israeli cultivar, Ananas Yok'neam. The RIL population was genotyped, and data was used to construct an ultra-dense, high-quality genetic map (N=5,663 binned SNPs) anchored to the *C. melo* genome. The utility of the densely genotyped population was demonstrated through QTL mapping of resistance to Fusarium wilt caused by *Fusarium oxysporum* f. sp. *melonis* race 1, a trait our laboratory and others have reported on previously. A major QTL co-located with the previously validated resistance gene *Fom-2*. In addition, three minor QTL and an epistatic interaction contributing to *Fom* race 1 resistance were identified. The MR-1 x AY RIL population provides a valuable resource for future QTL mapping studies and marker-assisted selection of disease resistance in melon.

P0722: Fruit Species

Phylogenetic Analysis of an Orphan Melon, *Cucumis melo* ssp. *agrestis* var. *texanus via* Genotyping-by-Sequence

Kaori Ando, William M. Wintermantel and James D. McCreight, USDA-ARS, Crop Improvement and Protection Research, Salinas, CA

Cucumis melo ssp. *agrestis* var. *texanus* Naudin is a wild/feral melon distributed throughout the southeastern U. S. from Florida to Texas, and in eastern and western Mexico. It was considered an escape of ssp. *melo* var. *chito* or var. *dudaim* until designated as a separate *varietas* of melon, though its origin remains, however, largely unknown. It is weedy and bears prolific, small round fruit, but has exhibited genetic potential for powdery mildew resistance for introgression to cultivated melon. All 44 accessions of var. *texanus* in the U.S. National Plant Genetic Resources System and 24 *C. melo* accessions from 12 other melon *varietas* in the two melon subspecies, *melo* and *agrestis,* were genotyped using the Genotyping-by-Sequencing method. Var. *texanus* accessions formed their own clade in phylogenetic analysis*,* distant from var. *chito* (1 accession) and var. *dudaim* (3 accessions), which also formed a clade distinct from the other with 20 representatives tested of 10 other melon *varietas*. Principal component analysis also placed var. *texanus* in a unique cluster distinct from two other melon clusters. Previous research revealed Cucumis melo endornavirus (CmEV) in all 25 representatives sampled from 12 of the 17 *varietas* of the two *C. melo* subspecies, and the analysis suggested the virus co-diverged with both subspecies and their *varietas*. In contrast, only 7% (3/44) of the var. *texanus* accessions were positive for the presence of CmEV, which suggests a unique origin or early isolation for this orphaned *varietas*.

P0723: Fruit Species

QTL Mapping of Sulfur Tolerance in Melon

Sandra E. Branham[1], James Daley[2], Melanie Katawczik[1], Amnon Levi[1] and Patrick Wechter[1], (1)USDA-ARS, Charleston, SC, (2)Clemson University, Clemson, SC

Elemental sulfur is a cheap, effective fungicide with multi-site action, which inhibits the evolution of pathogen resistance. Fungal pathogens cause significant yield losses in melon production. Many melon genotypes, however, suffer leaf necrosis in response to elemental sulfur application preventing use of this potent fungicide. Melon breeding efforts would benefit from incorporation of sulfur tolerance but little is known about its genetic basis. Here we present the results of a QTL mapping study in a densely genotyped *Cucumis melo* RIL population segregating for sulfur tolerance. Multiple QTL mapping identified two QTL and one epistatic interaction that explained 59.8% of the population variation for sulfur tolerance.

P0724: Fruit Species

Validation and Fine-Mapping of a Major Flowering Time Locus *Qdff3-1* in Watermelon

Winnie Gimode, Josh Clevenger and Cecilia McGregor, University of Georgia, Athens, GA

Watermelon (*Citrullus lanatus*) is a major vegetable crop grown worldwide mainly for its sweet, juicy fruit. Production is hindered by biotic and abiotic stresses, which are intensified by long production cycles. Flowering time is crucial in watermelon breeding as it is a major determinant of earliness, which dictates time of fruit set. Seedless watermelon production also relies on synchronized flowering between pollenizers and triploid cultivars. Developing high throughput molecular markers applicable in marker assisted selection (MAS) of flowering time in watermelon would potentially aid in selection for the early flowering trait, which would shorten the production time. Seedless watermelon breeding could also be enhanced through selection of the most suitable pollenizers for the triploids. A major quantitative trait locus (*Qdff3-1*: 12Mbp-17Mbp) responsible for ≈50% of the phenotypic variation observed for days to female flower was previously identified on chromosome 3 of watermelon. Potential candidate genes underlying the locus include *FT, TEMPRANILLO* and a recently identified PIP-kinase (*Cla002795*). The objective of this study was to validate and fine-map the *Qdff3-1* locus using single nucleotide polymorphism (SNP) markers. QTL-seq and candidate gene sequencing were employed to identify SNPs associated with the flowering trait. Markers were tested in a recombinant inbred line mapping population and validated on a cultivar panel to establish marker-trait association and determine their applicability in MAS of flowering time in watermelon. SNPs that will be potentially useful for the refinement of the QTL have been identified and may be applicable in MAS of flowering time in watermelon.

P0725: Fruit Species

Time-Series Transcriptome Analysis to Identify Resistance Mechanism of Anthracnose caused by *Colletotrichum orbiculare* Race 1 in Watermelon

Yoon-Jeong Jang[1], Minseok Seo[2], Bal-Kum Han[1], Sun-Ju Rhee[3] and **Gung-Pyo Lee**[1], (1)Dept of Integrative Plant Science, Chung-Ang University, Ansung, South Korea, (2)Department of Medicine, Harvard Medical School, (3)Department of Plant Sciences, The University of Cambridge, United Kingdom

Anthracnose caused by *Colletotrichum orbiculare* is one of the representative diseases that directly affect seedling survival and fruit harvests of watermelons. Here, we attempt to identify differentially expressed genes (DEGs) between resistant (DrHs7250) and susceptible (Oto9491) breeding lines to the anthracnose over time after *C. orbiculare* inoculation using time-series RNA-sequencing data. The expression levels of 20137 genes were measured in all 24 samples using the reference based RNA-seq pipeline, which quantifies gene expression levels based on reference genome, and associated gene model derived from Cucurbit Genome Database. In the time-series transcriptome analysis, 2102 genes had evidence of interaction, at FDR adjusted $p < 0.01$. In the post-hoc analysis, 49, 129, and 1796 genes were found to be significantly different between resistant and susceptible cultivars on 1, 3, and 5 days, respectively. Of the 1796 genes, 798 genes were selected up-regulated in DrHs7250 at that time and this set of genes was significantly enriched for functional annotations, including plant defense, Tify domain7, Jasmonic acid signaling pathway, regulation of jasmonic acid mediated signaling pathway , and Ethylene signaling pathway . From these functional terms, ERF gene-family and JAZ gene-family, were commonly identified and enriched. These results suggest that the jasmonic acid/ethylene signaling pathways may play a role in the protective mechanism conferred by *Colletotrichum*. We expect that this study contributes a deep understanding of defense mechanisms and plant/cell death processes at molecular levels by profiling time-series gene expression levels that varies with anthracnose disease progression.

P0726: Fruit Species

Assessment of Genetic Diversity for Kiwifruit (*Actinidia spp.*) using SNPs Developed by Genotyping-by-Sequencing

Daeil Kim[1], Sewon Oh[1,2], Keumsun Kim[1,2], Hyeondae Han[1,2], Kidong Hwang[1,2], Min-Gi Lee[1,2], Suji Ahn[1,2], Sunheum Cho[1,2], Mockhee Lee[3] and yong-Bum Kwack[3], (1)Chungbuk National University, Cheongju, South Korea, (2)Brain Korea 21 Center for Bio-Resource Chungbuk National University, Cheongju, South Korea, (3)Namhae Branch, National Institute of Horticultural and Herbal Science, Namhae, South Korea

To investigate genetic diversity in kiwifruits (*Actinidia* spp.), genotyping-by-sequencing (GBS) was applied. Using single nucleotide polymorphisms detected by GBS, phylogenetic relationship and population structure were analyzed in 89 kiwifruits accessions. The 89 kiwifruits accessions were clearly divided into two groups and the groups were characterized by presence and absence of hairs on pericarp. As a result of population structure analysis, the peak of delta *K* was detected at $K = 5$. Each cluster represented *A. arguta*, *A. chinensis*, *A. deliciosa*, *A.* hybrid, and wild accessions. In order to provide useful information to germplasm collection, the genetic relatedness and population structure were also investigated within *A. arguta* accessions. In phylogenetic tree, the 39 *A. arguta* accessions were characterized by male and female accessions, respectively. As a result of principal component analysis, three genetic population were observed and there were two genetic populations in the female *A. arguta* accessions. These results provide an information to develop a core germplasm set for *Actinidia* spp. Moreover, additional data such as agronomical traits may be needed to classify the *Actinidia* accessions in detail.

P0727: Fruit Species

Comparative Genomics of the Juglandacea

Alexander Trouern-Trend, Sumaira Zaman and Jill L. Wegrzyn, Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

The *Juglandaceae* (or walnut family) represents 50 species of deciduous aromatic trees, of which 17 reside in North America. There are eight genera in this family, including *Juglans* and *Pterocarya*, which represent many of the commercially valuable nut and wood-producing trees. Although *Juglans regia* is the predominant nut producer, it is often grafted onto other members of Juglandaceae to improve the production, pest resistance, and drought tolerance in orchards. Some of these species are threatened in their native range and pose a direct threat to the cultivation.

The analysis of of *J. regia* (Persian walnut), *J. microcarpa* (Texas Walnut), *J. cathayensis* (Chinese Walnut), *J. hindsii* (Hinds' Black Walnut), and an outgroup that belongs to the same family as the walnut (*Juglandaceae*), *Pterocarya stenoptera* (Chinese Wingnut) provides meaningful insight to accelerate breeding and facilitate genetic dissection of complex traits for preservation. Here, we detail the process of annotation and comparative analysis of the gene space in these important species. The softmasked genomes were processed through an annotation pipeline that utilized BRAKER *ab initio* gene prediction software. BRAKER leveraged transcriptomic data from 22 *J. regia* RNA-Seq libraries from various tissue types and developmental stages. Additional evidence and filters were applied to reinforce the evidence for and filter multiexonic and monoexonic putative genes predicted by BRAKER to yield between 32,000 and 40,000 genes. The final high quality annotations were processed to examine comparative genomics across these five important species.

P0728: Fruit Species

Transcriptomic Analysis of Flower Initiation Genes in Pecan

Hormat Shadgou[1], Spencer D. Barnes[1], Jeffery Fisk[2], Joe Song[1], Richard Heerema[1] and Jennifer J. Randall[1], (1)New Mexico State University, Las Cruces, NM, (2)Yale University, New Haven, CT

Carya illinoinensis (pecan) is an economically important nut tree species. Pecan is native to North America and to date is cultivated in 27 states in United States. Pecan has also been introduced to various environments around the world. Alternate bearing, defined as low production of high quality nuts in one year and high production of lower quality nuts the next year, is a major constraint to reliable pecan production. The genetic information of the flower initiation timing in pecan is unknown. The hypothesis of this study is that pistillate flower initiation occurs in two separate steps. Based on transcriptomic data from a time course study, there is evidence that the first step may begin a year before the appearance of flowers on the tree and continue after dormancy is over in the spring. RNASeq analysis of bud tissues from 'Western' cultivar pecan trees elucidated differential gene expression in more than 40 genes during the growing season. Moreover, flowering genes showed different levels of expression between the flowering and non flowering samples. This primary DESeq provided the first glimpse of the differentially expressed genes in 'Western' pecan buds during the period from June to September. It also provided gene networks and an estimate timing for the initiation of pistillate flowering. The results of this study were used to design the next step through characterizing the responsible signals for flower induction, the time of signaling and the process behind flower initiation in pecan.

P0729: Fruit Species

Molecular Mechanisms Regulating Nut Weight in *Castanea crenata*

Eung-Jun PARK, Hyoshin Lee and Hyo-Ryeon Lee, National Institute of Forest Science, Republic of Korea, Suwon, South Korea

Chestnut (*Castanea crenata*) is an important short-term income tree crop in Republic of Korea. Traditional selections on chestnut quality resulted in a huge diversity of both weight and sugar content, which further provide excellent resources to exploit the molecular bases of this diversity. To systematically understand its molecular mechanisms during fruit development, we first performed both metabolomic and transcriptomic analyses with two small (JW) and large (DH) nut-bearing varieties. A total of 42 water-soluble metabolites were differentially accumulated in nut tissues between JW and DH. Among those metabolites, the contents of monosaccharides were significantly different between two varieties during the entire periods of fruit development. Interestingly, the content of sucrose in both leaf and fruit tissues at Stage III showed significant correlations with nut weights ($r>0.7$, $p<0.05$). We also generated 37,649 unigenes matching 97.9% of the *C. mollissima* genome using single molecule real-time (SMRT) sequencing. Further transcriptome analysis using Illumina high-throughput paired-end sequencing provided a number of differentially expressed genes (DEGs) involved in seed development and fruit size, respectively.

P0730: Fruit Species

Construction of a Genetic Map for Longan (*Dimocarpus longan*) using RAD-Seq

Julie K. Nguyen[1], Shaoquan Zheng[2], Jianguo Li[3], Ching Man Wai[1] and Ray Ming[1,4], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)Fujian Agriculture and Forestry University, Fuzhou, China, (3)South China Agricultural University, Guangzhou, China, (4)FAFU and UIUC-SIB Center for Genomics and Biotechnology, Fuzhou, China

Longan (*Dimocarpus longan* Lour.) is an economically important tropical crop in southeast Asia. A high-density genetic map is essential for genome anchoring and mapping traits, however, genetic resources are limited for this species. Therefore, we developed a SNP-based genetic map using double-digest restriction site-associated DNA sequencing (RAD-seq). We sequenced RAD-seq libraries of parental cultivars 'Lidongben', 'Qingkeyuanbao' and their 63 F1 progenies using an Illumina platform, and the reads were aligned to the published 'Honghezi' reference genome. The maternal 'Lidongben' map was constructed with 1025 markers, containing 15 linkage groups with a total distance of 4096.4 cM and with an average distance of 4.27 cM per marker. For the paternal 'Qingkeyuanbao' map, it contained 750 markers with the same number of linkage groups, spanning 2619.4 cM and an average marker interval distance of 3.72 cM. Using these two maps, we anchored 381 scaffolds, comprising 50.5% (224.7 Mb) of the published genome. Breeding efforts have been delayed by longan's high heterozygosity and long juvenile period. This genetic map can aid in marker-assisted selection and identify QTLs of agronomically important traits.

P0731: Fruit Species

Determination of Morphological, Anatomical, Physiological and Gene Expression Changes of some Olive Cultivars under Water Deficient Stress

Zerbab Guler[1], Mucahit Taha Ozkaya[1], Canan Yuksel Ozmen[2], Ali□ Ergul[2] and Basak Ozdemi□r[1], (1)Ankara University, Ankara, Turkey, (2)Ankara University Biotechnology Institute, Ankara, Turkey

Olive (*Olea europaea* L.) is one of the most significant fruit-tree crop species of the Mediterranean region. Dehydrins which are expression may be induced by developmental process and environmental stresses that lead to cell dehydration; have important functions in response to various stress factors. Although genomes of olive have been published, there is a little information about Dehydrins in this species. In this study, a 90 -day long drought experiment was conducted using two-year old plants of four olive cultivars (Manzalina, Ayvalik, Domat and Kilis Yaglik). For this purpose, olive nursery plants exposed to water stress for 45 days followed by 45 days of post-stress period, in a growth chamber. In order to determine the gene expression changes, total RNA was extracted from leaf samples after 15, 45 and 90 days of drought-treatment as well as after 15, 30 and 45 days of post-treatment period and Dehydrin genes were comparatively analyzed by q-RT PCR. Furthermore, leaf relative water content, leaf water potential were studied.

Key Words: *Olea europaea* L., water deficient stress, morphology, anatomy, physiology, Dehydrin(DHN) genes

P0732: Fruit Species

Whole-Genome Assembly of *Cocos nucifera* Var. Laguna Tall (LAGT) with Short and Long Reads Improves Contiguity and Gene Space Completeness

Cynthia P. Saloma[1,2], Charles Anthon E. Cadorna[2], Angelica Khryss Yvanne C. Ladisla[2], Ma. Regina G. Punzalan[2] and Zenaida Magbanua[2], (1)National Institute of Molecular Biology and Biotechnology, University of the Philippines, Quezon City, Philippines, (2)Philippine Genome Center, University of the Philippines, Quezon City, Philippines

Cocos nucifera (coconut) that belongs to the Arecaceae family and the only reported species under the genus cocos is cultivated in the Philippines due to its extensive application in agriculture and industry. Here, we report our analyses on the draft genome assembly of *Cocos nucifera* var. Laguna Tall (LAGT). We have generated paired-end libraries and sequenced using Illumina NextSeq500. The output raw reads were pretreated before assembly. A reference-based approach was used to assemble the *Cocos nucifera* var LAGT genome using the software SOAPaligner. Then, gaps within scaffolds were filled by GMcloser with Moleculo TruSeq synthetic long reads. Assessment of genome contiguity was done using QUAST. Based on QUAST results, the N50 of the assembly is approximately 130,790 bp and contains 60,745 scaffolds spanning ~81% of the genome. In evaluating gene space completeness, we used the BUSCO pipeline (evolutionary-informed expectations of gene content from near-universal single-copy orthologs selected from OrthoDB), and BLAT alignment of unigenes (transcript obtained from different tissues of coconut) to the assembled genome. BUSCO analysis showed that there are about 92.5% of 956 expected plant genes were identified as complete. BLAT alignment results indicated that the assembled genome covered 98.6% of the expressed unigenes. This suggests a high level of gene coverage for the assembled genome.

P0733: Fruit Species

Coconut Genetics and Genomics for Host Insect Resistance

Hayde F. Galvez[1], Darlon V. Lantican[1], Maria Luz J. Sison[1], Roanne R. Gardoce[1], Barbara L. Caoili[1], Alma O. Canama[1], Melvin P. Dancel[1], Romnick A. Latina[1], Cris Q. Cortaga[1], Don Serville R. Reynoso[1], Michelle S. Guerrero[1], Susan M. Rivera[2], Ernesto E. Emmanuel[2], Cristeta A. Cueto[2], Susan R. Strickler[3] and Lukas A. Mueller[3], (1)University of the Philippines Los Baños, College, Laguna, Philippines, (2)Philippine Coconut Authority (PCA), Quezon City, Philippines, (3)Boyce Thompson Institute, Ithaca, NY
The Philippines is the second world supplier of coconut (*Cocos nucifera* L.) by-products. Used to be the topmost producer/supplier, the country however has been threatened with serious production constraints including the recent outbreak of coconut scale insect (CSI) in major coconut regions. To facilitate the development of insect resistant variety, the advancement in genomics and related technologies are harnessed for their optimum integration in the coconut breeding program. Coconut NGS reads were generated by sequencing the whole gDNA of Catigan Green Dwarf (CATD) coconut variety using several sequencing platforms i.e. 50X Illumina Miseq, 15X PacBio SMRT, and Dovetail Chicago sequencing. Based on combined analysis, a total genome sequence length of 2.1 Gb consisting of 8,062 scaffolds with N50 value of 569 kb was assembled. The genome assembly and initial gene models were uploaded in a local genome database and utilized to develop DNA markers targeting candidate genes for insect resistance. The genome sequence of CSI was also characterized and a species-specific DNA marker system was developed. Employing both 'Choice' and 'No-Choice' tests, 73 core coconut germplasm and on-farm outstanding selections were assayed for host resistance against CSI. The coconut materials were also mined for point-mutations at the candidate gene sequences, which are being associated to the differential CSI host response. RNA-sequence data from these differential gene expressions are also being analyzed, output of which will be mapped back to the coconut genome for a targeted functional annotation. Through forward/reverse genetics correlation studies coupled with relevant genome and bioinformatics data, the mechanism of host insect resistance will be elucidated in selected coconut varieties.
Keywords: *Cocos nucifera* L., coconut scale insect, next-generation sequencing, host insect resistance, molecular marker, EcoTILLING, genomics-assisted breeding

P0734: Fruit Species

Genetic Diversity Analysis and EcoTILLING in Coconut (*Cocos nucifera* L.) for Host Insect Resistance

Melvin P. Dancel, Rei Jediael O. Guevarra and **Hayde F. Galvez**, University of the Philippines Los Baños, College, Laguna, Philippines

Coconut (*Cocos nucifera* L.), also known as the 'Tree of Life', is one of the most important crop in the world. The study is aimed to analyze the genetic diversity using 16 microsatellite markers (SSR) mapped from each coconut chromosome and to mine natural allele variants of identified plant glandular trichomes and host insect resistance-associated genes across 73 varieties of coconut that are conserved at the Philippine Coconut Authority (PCA) genebank. The overall outcome of the genetic cluster analysis supports the distinct ecotype classification of coconut based on probable area of origin - the Indo-Atlantic and the Indo-Pacific coconuts; as well as the country/region of popular cultivation and/or germplasm source. Results from the study also support the concept that Dwarf varieties were domesticated from Tall population, but whether the Pacific Region was indeed the center of coconut domestication has remained for validation using more in-depth molecular, morphology and socio-cultural evidences. On the other hand, PCR conditions were optimized to mine the natural variants of identified genes associated to glandular trichomes and host insect resistance across coconut varieties in the Philippines via Targeting Induced Local Lesions In Genomes (EcoTILLING). Natural SNPs on coronatine-insensitive 1 (COI 1) and polyphenol oxidase (PPO) loci are screened in different coconut varieties. The SNPs will be further characterized for high-throughput screening and selection of favorable alleles in genomics-assisted coconut breeding for outstanding coconut varieties.

P0735: Fruit Species

The Coconut Genome: Providing a Reference Sequence Towards Coconut Varietal Improvement

Darlon V. Lantican[1], Susan R. Strickler[2], Alma O. Canama[1], Roanne R. Gardoce[1], Lukas Mueller[2] and Hayde F. Galvez[1], (1)University of the Philippines Los Baños, College, Laguna, Philippines, (2)Boyce Thompson Institute, Ithaca, NY

We present the whole genome sequencing (WGS) of Catigan Dwarf (CATD) coconut variety, chosen for its genome simplicity and low heterozygozity. PacBio SMRT sequence data was generated at 15X coverage and corrected with assembled 50X Illumina paired-end MiSeq reads. Through the hybrid assembly approach, the draft assembly of the dwarf coconut genome has N50 of 120 kb. The genome was further improved through Dovetail Chicago sequencing. The input de novo hybrid assembly, Miseq PE reads and Chicago library reads were used as input data in the HiRise pipeline in order to scaffold the genome assembly. As a result, the final assembly has now a total sequence length of 2.1 Gb consisting of 8,062 scaffolds with N50 value of 569 kb. This covers around 97.6% of the estimated CATD genome of 2.15 Gb based on the homozygous k-mer peak. Around 1.556 Gb is identified as interspersed repeat sequences or 73.93% of the total assembled genome. A total of 35,231 high-confidence gene models are identified using the MAKER annotation pipeline. Result of the BUSCO analysis has revealed a 83.1% completeness of the current genome annotation, and 7.4% fragmented single copy orthologs (USCO) based on 1440 genes in the plant specific OrthoDB database. The assembly statistics and quality evaluation results demonstrate that our current draft scaffold assembly has covered most of the dwarf coconut and gene units. Such provides a good reference coconut genome for various applications such as re-sequencing projects, gene mining and DNA marker development, and functional annotation/genomics approaches.

P0736: Legumes, Soybean, Common Bean, and related

Genomic Resources for Soybean Research at NCBI

Anjana Raina Vatsan, Francoise Thibaud-Nissen and Terence D. Murphy, National Center for Biotechnology Information (NCBI), Bethesda, MD

The National Center for Biotechnology Information (NCBI) provides a range of resources and tools for storage and analysis of genomic data from a wide variety of organisms. A diverse subset of genome assemblies publicly available in the International Nucleotide Sequence Databases (GenBank/DDBJ/EMBL) are selected for annotation by NCBI's Eukaryotic Genome Annotation Pipeline and inclusion in NCBI's RefSeq dataset (https://www.ncbi.nlm.nih.gov/refseq/). To date, 73 plant genomes have been annotated by NCBI, including 8 species in the Fabaceae family. Three plant species, soybean, corn, and tomato, have been selected for further expert curation by the NCBI RefSeq group. Expert curation ensures accurate and full length representation of nucleotide and protein sequences and to resolve data conflicts and

ambiguities. Gene and protein names are assigned and publications added, when available. Gene-specific data is available in NCBI's Gene resource (https://www.ncbi.nlm.nih.gov/gene/), and gene annotation and other data can be explored in NCBI's Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/). RefSeq data can be accessed from the RefSeq homepage https://www.ncbi.nlm.nih.gov/refseq/ or can be downloaded from the FTP directory at ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/plant/Glycine_max/.

P0737: Legumes, Soybean, Common Bean, and related

KnowPulse: An Integrated Bioinformatic Web Portal for Pulse Crops Providing Genomic, Genotypic and Phenotypic Data and Breeder-Focused Analysis Tools

Lacey-Anne Sanderson, Reynold Tan, **Carolyn T. Caron**, Crystal Chan and Kirstin Bett, University of Saskatchewan, Saskatoon, SK, Canada

With the deluge of genotypic and phenotypic data being generated, it is essential to ensure access to up-to-date resources and analysis tools for breeders to maximize genetic gain. We have added new and exciting features to KnowPulse (http://knowpulse.usask.ca), a web-based resource for plant breeders and geneticists interested in pulse crops. In addition to serving as the digital home for the pulse breeding program at the University of Saskatchewan with project lists, released varieties and publications, KnowPulse provides access to genomic sequence, transcript sequences, variants and markers for chickpea, common bean, field pea and lentil. Genotypic data is summarized on marker pages and detailed in a researcher-filtered marker-by-germplasm table for comparison between germplasm. Genomic data can be utilized via crop-specific JBrowse instances and BLAST databases. Current progress is underway to provide access to phenotypic data via trait and germplasm pages, including trait-specific data visualizations. Additionally, we plan to integrate KnowPulse with other legume databases through the Legume Federation to allow two-way sharing of data. These tools allow breeders to enhance crossing design, improve germplasm development and associate marker genotypes with current breeding materials.

P0738: Legumes, Soybean, Common Bean, and related

legumeinfo.org: Legume Research and Trait Data that is Findable, Accessible, Interoperable and Re-Usable

Steven B. Cannon[1], Jacqueline D. Campbell[2], Sam Hokin[3], Joel Berendzen[3], Ethalinda Cannon[2], Alan M. Cleary[4], Sudhansu Dash[3], Connor T. Cameron[3], Wei Huang[2], Scott R. Kalberer[1], Amanda Cooksey[5], Nathan Weeks[1] and Andrew Farmer[3], (1)USDA-ARS-CICGRU, Ames, IA, (2)Iowa State University, Ames, IA, (3)National Center for Genome Resources (NCGR), Santa Fe, NM, (4)Montana State University, Bozeman, MT, (5)University of Arizona, Tucson, AZ

The Legume Information System (LIS; https://legumeinfo.org) is a resource for comparative legume research. Currently, LIS hosts annotated genomes of 14 species representing both crop and model legumes. LIS partners with other plant data resources in the "Legume Federation" (eg. jcvi.org/medicago, SoyBase, PeanutBase and Phytozome) to integrate genomic, genetic, trait and other data, and to provide links to other legume data sets not housed at LIS. Through synteny and phylogenetic-based methods, LIS uses data from well-studied legume species (e.g. soybean and Medicago) to facilitate comparative research across species. Gene families allow the researcher to traverse paralogous and orthologous sequences across legume species. In the past year, LIS has seen the following improvements, in coordination with the Legume Federation project: a new comparative genetic map viewer (CMap-js), several new legume "mines" (instances of the InterMine biological data warehouse, http://intermine.org), new legume-focused gene families and tools for placing user-submitted genes into the families and trees, and a new cowpea genome and gene sets. A major goal of LIS is to assist researchers in discovering important trait, genomic, and genetic information in the vast amount of data available today. LIS is a collaborative project of the USDA-ARS, NCGR, and the many dedicated researchers working on legume biology.

P0739: Legumes, Soybean, Common Bean, and related

The Federated Plant Database Initiative for the Legumes

Andrew Farmer[1], Jacqueline D. Campbell[2], Sudhansu Dash[1], Wei Huang[2], Malachy O'Connell[2], Andrew Wilkey[3], Akshay Yadav[2], Nathan T. Weeks[3], Joel Berendzen[1], Alan Cleary[4], Sam Hokin[1], Connor T. Cameron[1], Vivek Krishnakumar[5], Amanda Cooksey[6], Agnes P. Chan[5], Ethalinda Cannon[2], Eric Lyons[7], Christopher D. Town[5], Steven B. Cannon[3] and David Fernandez-Baca[2], (1)National Center for Genome Resources (NCGR), Santa Fe, NM, (2)Iowa State University, Ames, IA, (3)USDA-ARS-CICGRU, Ames, IA, (4)Montana State University, Bozeman, MT, (5)J. Craig Venter Institute, Rockville, MD, (6)University of Arizona, Tucson, AZ, (7)University of Arizona; BIO5 Institute; CyVerse, Tucson, AZ

The "Legume Federation" (https://legumefederation.org) is an NSF project to facilitate use of data standards, distributed development, and comparative analysis, to support research across the legume family – and to support robust agriculture for a world that is significantly legume-fed. Tools and resources added in the last year include a new interactive genetic map viewer, CMap-js, additions to the Data Store for major legume data sets, improvements to the Genomic Context Viewer (for exploring genome- and gene-scale synteny between legume species), new legume-focused gene families and phylogenetic trees, a tool for placing user-submitted gene sequences into gene families and trees, and a new InterMine instance for cowpea - bringing the collection of legume "mines" to seven (bean, chickpea, cowpea, soybean, peanut, Medicago, and cross-legume). Participating Genomic Data Portals (GDPs) currently include, but are not limited to MedicagoGenome (http://medicagogenome.org), SoyBase (https://soybase.org), PeanutBase (https://peanutbase.org), the Legume Information System (https://legumeinfo.org), CyVerse (http://www.cyverse.org), Phytozome (https://phytozome.jgi.doe.gov), Climate Resilient Chickpea Lab (http://chickpealab.ucdavis.edu), NSF Cowpea Genome Project (https://www.nsf.gov/awardsearch/showAward?AWD_ID=1543963), Alfalfa Breeder's Toolbox (https://www.alfalfatoolbox.org/), Medicago Hapmap project (http://www.medicagohapmap.org), KnowPulse (http://knowpulse.usask.ca), and the Cool Season Food Legume Database (http://www.coolseasonfoodlegume.org).

P0740: Legumes, Soybean, Common Bean, and related

What's New at SoyBase

David Grant[1], Rex Nelson[2], Kevin H. Feeley[2], Shawn Conners[2], Victoria Carollo Blake[3], Jacqueline D. Campbell[4], Wei Huang[4], Nathan Weeks[2] and Steven B. Cannon[2], (1)USDA-ARS-CICGR & Iowa State University, Ames, IA, (2)USDA-ARS-CICGRU, Ames, IA, (3)USDA ARS WRRC, Albany, CA, (4)Iowa State University, Ames, IA
SoyBase, the USDA-ARS soybean genetics and genomics database, provides a comprehensive collection of data, analysis tools and links to external resources of interest to soybean researchers. SoyBase is an actively curated database with new data regularly being incorporated, including updates to the controlled vocabularies (ontologies) for soybean growth, development and phenotypic traits, soybean genes, QTL, and genome sequences and annotations. Some notable recent additions include a pedigree search tool, a search tool for GRIN descriptor data, and >100 gene expression studies. The data in SoyBase are provided through intuitive interfaces, and are linked together wherever possible to allow easy identification and browsing of related subjects.
The SoyBase home page contains the SoyBase Toolbox, which provides quick access to a search of SoyBase, the SoyCyc metabolic pathways, a comprehensive data download page, a genome BLAST tool, and direct links to the genetic maps and genome sequence browser. An extensive navigation menu and site description for all sections of SoyBase is provided. Searching at SoyBase uses an underlying trait-based approach to return all information related to the search term. Numerous data types are available including genetic and QTL maps, the reference genome sequence with annotation tracks covering genetic markers, genome organization, and gene annotation and expression. Pedigrees for entries in the Soybean Uniform Trials are available. SoyBase also includes an extensive genechip and RNA-Seq gene expression atlas, innovative tools for identifying fast neutron-induced mutants affecting genes or traits of interest, and several 'omics tools, for example a GO Term Enrichment tool, that enable sophisticated queries on lists of genes.

P0741: Legumes, Soybean, Common Bean, and related
Targeted Next Generation Sequencing Approaches in Corn, Cucumber and Soy for High Throughput Genotyping
Michelle Swimley[1], Christina Buchanan-Wright[1], Angela Burrell[1], **Michelle Swimley**[1], Roy C. Willis[1], Jie Lu[1], Livio Baselgia[2], Haktan Suren[1] and Rick Conrad[1], (1)Thermo Fisher Scientific, Austin, TX, (2)Thermo Fisher Scientific, Schlieren, Switzerland
With advances in plant phenotyping approaches for quantitative genetic analysis and increasing complexity of gene pyramiding schemes, the number of markers required for successful molecular breeding programs in agriculture is increasing. Historically, technology has been polarized between high marker, high cost microarrays or low cost singleplex approaches that are not easily scalable. Targeted genotyping by sequencing (GBS) is emerging as a powerful alternative for mid-density genotyping of 100s to thousands of marker in a high throughput and cost-effective manner.
We have applied AgriSeq targeted GBS, a high throughput amplicon based sequencing workflow performed on the Ion S5 system, to three economically important crops. A 2800 marker cucumber panel, an 1100 marker soy panel and two independent corn panels targeting 900 and 1000 markers were designed for the AgriSeq workflow. The average genotyping marker call rate ranged between 91%-98% for these panels, with >94% average uniformity and >99% on-target reads. >99.4% reproducibility has been demonstrated for this workflow over multiple independent library preparations and sequencing runs, with genotype concordance to orthogonal array technologies of 99.4%. Compatible with high throughput processing, the AgriSeq approach can multiplex up to 768 samples simultaneously and generate up to 1.6M genotypes per day. In addition, this approach allows for the discovery of additional SNPs and micro-haplotypes around the targeted markers which enable further traceability and association in downstream studies. These results demonstrate the utility of the AgriSeq targeted Genotyping by Sequencing for plant molecular breeding programs.

P0742: Legumes, Soybean, Common Bean, and related
Structural Variation Detection using SMRT Sequencing
Gregory T Concepcion[1], Shreyasee Chakraborty[1], Michelle Vierra[1], Pradeep Reddy Marri[2] and Aaron Wenger[1], (1)PacBio, Menlo Park, CA, (2)Dow AgroSciences, Indianapolis, IN
Structural variants (genomic differences ≥50 base pairs) contribute to the evolution of traits and disease. Most structural variants (SVs) are too small to detect with array comparative genomic hybridization and too large to reliably discover with short-read DNA sequencing. While *de novo* assembly is the most comprehensive way to identify variants in a genome, recent studies in human genomes show that PacBio SMRT Sequencing sensitively detects structural variants at low coverage. Here we present an example of SV characterization in two major crop species grown world-wide, Zea mays (Maize) and Glycine max (Soy). Structural Variants were called for both species from both long read and short read datasets using PacBio's PBSV variant calling pipeline and Illumina's Manta variant calling pipeline respectively. PBSV was able detect up to 22 times as many variants in Soy, and 5 times as many variants in Maize, indicating that despite a coverage disadvantage, the long read datasets were responsible for identifying substantially more SV locations in the respective genomes analyzed. This study demonstrates the feasibility of a low-coverage approach to high-throughput SV discovery in agriculture.

P0743: Legumes, Soybean, Common Bean, and related
A High-Quality Wild Soybean Reference Genome
Min Xie, Claire Yik-Lok Chung, Ting-Fung Chan and Hon-Ming Lam, Center for Soybean Research of the Partner SKL of Agrobiotechnology, and School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, Hong Kong
Wild soybean (*Glycine soja*) contains high genetic diversity and could serve as a natural gene pool for crop improvement programs of soybean. To better understand the genetic blueprint of wild soybean and extract useful genes and alleles therein, a fine genome is imperative. However, only draft genomes are available for wild soybean to date.
Here we report a near complete genome assembly for a wild soybean accession (W05), constructed via *de novo* assembly using a combination of PacBio single-molecule sequencing, Illumina paired-end sequencing, Bionano Genomics optical mapping, and Dovetail Genomics Hi-C technologies. The final assembly is 1.01 Gb in length, with 95.7 % of sequences anchored to 20 chromosomes. Contig N50 of this assembly is 3.33 Mb, representing a 17X improvement over the current reference assembly for cultivated soybean (*Glycine max var. Williams 82* v275). The collinearity between chromosomes of the wild soybean W05 and the cultivated soybean William 82 is very high, except a previously reported reciprocal translocation between chromosome 11 and chromosome 13 found in the W05 genome.To facilitate genome annotation, we

sequenced 1.3 billion paired-end Illumina RNA-seq reads for various tissues of various development stages. In addition, we also sequenced 827,560 PacBio IsoSeq reads for mixed RNA samples from multiple tissues.
The wild soybean genome resources will benefit soybean research community and breeders all around the world.

P0744: Legumes, Soybean, Common Bean, and related
Assembly and Annotation of a Draft Genome Sequence for *Glycine latifolia*, a Perennial Wild Relative of Soybean
Qiong Liu, University of Illinois, Urbana, IL and Leslie Domier, USDA-ARS, Urbana, IL
Glycine latifolia (Benth.) Newell & Hymowitz (2*n*=40), as one of the 26 wild perennial relatives of soybean, possesses genetic diversity and agronomically favorable traits that are lacking in soybean. Whole genome sequence and annotation of *G. latifolia* will serve as a valuable source of alternative alleles and novel genes to facilitate soybean improvement. Here, we report the 939-Mb draft genome assembly of *G. latifolia* (PI 599298) using exclusively linked-reads sequenced from a single Chromium library. We organized scaffolds into 20 chromosome-scale pseudomolecules utilizing two genetic maps and the *Glycine max* genome sequence. Annotation of the assembled genome yielded 64,692 protein-coding loci. In comparative analysis with five legume species, genes related to defense responses were significantly overrepresented in *Glycine*-specific orthologous gene families. A total of 304 putative nucleotide-binding site (NBS)-leucine-rich-repeat (LRR) genes were identified in this genome assembly. Different from other legume species, we observed a scarcity of TIR-NBS-LRR (TNL) genes in *G. latifolia*. The *G. latifolia* genome was also predicted to contain genes encoding 367 LRR-receptor-like kinases (RLK), a family of proteins involved in basal defense responses and responses to abiotic stress. This genome assembly will be a valuable source of allelic diversity and novel genes to assist in genetic improvement on soybean. This study also highlights the efficacy and cost-effectiveness of the application of Chromium linked-reads in diploid plant genome *de novo* assembly.

P0745: Legumes, Soybean, Common Bean, and related
A Systematic Analytical Approach to Rapidly Identify Candidate Domestication-Related Genes
Davoud Torkamaneh, University of Guelph and Laval University, Quebec, QC, Canada, Jérôme Laroche, Laval University, Québec, QC, Canada and Francois Belzile, Laval University, Quebec, QC, Canada
Domestication is an important key co-evolutionary process through which humans have extensively altered the genomic make-up and appearance of both plants and animals. The identification of domestication-related genes remains very arduous. In this study, we present a systematic analytical approach that harnesses two recent advances in genomics, whole-genome sequencing and prediction of loss-of-function (LOF) mutations, to greatly facilitate the assembly of an exhaustive catalogue of domestication-related candidate genes. Using whole-genome sequencing data for 296 cultivated (*G. max*) and 64 wild soybean accessions, we identified 10,792 LOF variants, and 193 genes that are uniquely fixed for the LOF allele in domesticated soybeans. Existing soybean transcriptomic data led us to overcome analytical challenges associated with whole-genome duplications and to identify neo- or sub-functionalized genes. This systematic approach allowed us to identify 130 candidate domestication-related genes in an efficient and rapid way. This catalogue contains all of the previously well-characterized domestication genes in soybean, as well as some orthologues from other domesticated crop species. In addition, it comprises many additional novel candidate domestication genes. Overall, this collection of candidate domestication-related genes in soybean is almost twice as large as the sum of all previously reported candidate genes in all other crops. We believe this systematic approach could readily be used in wide range of species.

P0746: Legumes, Soybean, Common Bean, and related
Parallel Domestication of a Dormancy Gene in Crops from Multiple Families
Min Wang, Wenzhen Li, Chao Fang, Fan Xu, Yucheng Liu, Chengcai Chu and **Zhixi Tian**, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
Domesticated species often exhibit convergent phenotypic evolution, termed the domestication syndrome, of which loss of seed dormancy is a typical component. To date, the dormancy genes that contribute to parallel domestication have not been reported. Here, we cloned the classical stay-green *G* gene from soybean and surprisingly found that it controls seed dormancy and that it was strongly selected during soybean domestication. Moreover, its orthologs in rice and tomato also underwent selection during domestication. Transformation analysis confirmed that the *G* orthologs had conserved functions in controlling seed dormancy in soybean, rice and *Arabidopsis*. Further functional investigations demonstrated that G affected seed dormancy through interactions with NCED3 and PSY and in turn modulated abscisic acid synthesis. Therefore, we identified a gene responsible for seed dormancy, representing the first domestication gene that has been subject to parallel selection in multiple crop families and which may facilitate the domestication of new crops.

P0747: Legumes, Soybean, Common Bean, and related
Comparative Transcriptome Analysis of Flower Heterosis in Two Soybean F1 Hybrids by RNA-Seq
Chunbao Zhang and Limei Zhao, Jilin Academy of Agricultural Sciences, Changchun, China
Heterosis has been widely exploited as an approach to enhance crop traits during breeding. However, its underlying molecular genetic mechanisms remain unclear. Recent advances in RNA sequencing technology (RNA-seq) have provided an opportunity to conduct transcriptome profiling for heterosis studies. We used RNA-seq to analyze the flower transcriptomes of two F1 hybrid soybeans (HYBSOY-1 and HYBSOY-5) and their parents. More than 385 million high-quality reads were generated and aligned against the soybean reference genome. A total of 681 and 899 genes were identified as being differentially expressed between HYBSOY-1 and HYBSOY-5 and their parents, respectively. These differentially expressed genes (DEGs) were categorized into four major expression categories with 12 expression patterns. Furthermore, gene ontology (GO) term analysis showed that the DEGs were enriched in the categories metabolic process and catalytic activity, while Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis found that metabolic pathway and biosynthesis of secondary metabolites were enriched in the two F1 hybrids. Comparing the DEGs of the two F1 hybrids by GO term and KEGG pathway analyses identified 26 common DEGs that showed transgressive up-regulation, and which could be considered potential candidate genes for heterosis in soybean F1 hybrids. This identification of an extensive transcriptome dataset gives a comprehensive overview of the flower

transcriptomes in two F1 hybrids, and provides useful information for soybean hybrid breeding. These findings lay the foundation for future studies on molecular mechanisms underlying soybean heterosis.

P0748: Legumes, Soybean, Common Bean, and related

Assessment of Genetic Diversity and Historical Genomic Changes between Two Public Soybean Breeding Programs at the University of Guelph

Robert Bruce[1], Chris M Grainger[1], Milad Eskandari[2] and Istvan Rajcan[1], (1)University of Guelph, Guelph, ON, Canada, (2)University of Guelph-Ridgetown, Ridgetown, ON, Canada

To understand diversity within soybean breeding germplasm in Ontario, Canada, a 296 line panel named University of Guelph Germplasm Panel (UGGP) from two U. of Guelph soybean breeding programs was studied. The University of Guelph's Ridgetown campus breeding program (maturity group II) located in Ridgetown, Ontario contributed 34 released cultivars and 31 experimental lines. The Guelph campus breeding program (maturity groups 00, 0 and I) contributed 63 released cultivars and 76 experimental lines. The remaining genotypes represent the pedigrees of the Guelph campus breeding program to study selection signatures and historical trends in Ontario-adapted soybeans. Release dates for UGGP germplasm ranged from 1907 to 2016. The UGGP was field tested in 2015 and 2016 at two locations per breeding program in Ontario, Canada, to assess phenotypic diversity and study trends in trait improvement over year-of-release. Yield significantly increased in Group I soybeans at 18.8 $kg^{-1}ha^{-1}$year, while the overall efficiency of yield improvement was calculated at 0.11 $kg^{-1}ha^{-1}day^{-1}$year. Phenotypic data highlighted differing breeding objectives and regional adaptation between the breeding programs. UGGP lines were sequenced using genotyping-by-sequencing (GBS) and called using Fast-GBS resulting in 40,307 genome-wide single nucleotide polymorphisms (SNPs), with approximately 1 SNP per 23 kb. Linkage disequilibrium (LD) analysis of the UGGP lines indicated long regions of high LD within UGGP lines. Principal component and cluster analyses revealed overlap between germplasm from both breeding programs and historical germplasm. Understanding the genetic landscape of germplasm within a breeding program and the associated phenotypes will help breeders to make informed decisions to guide future breeding efforts.

P0749: Legumes, Soybean, Common Bean, and related

Leveraging Historic and Modern Soybean Uniform Regional Trial Results to Train Genomic Prediction Models

Benjamin W. Campbell[1], Peng Zhou[2], Liana Nice[1] and Aaron J. Lorenz[1], (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN

Genomic prediction is being utilized by plant breeders in multiple crop species as a method to accelerate genetic improvement. Initial examples of genomic prediction being used for crop improvement focused on large populations with simple family structures and relatively large amounts of phenotypic data per line. In reality, typical breeding programs utilize more complicated breeding designs with varying amounts of phenotypic data per line. In a typical breeding program pipeline, many lines from a family are screened on a limited basis early in development, and only a few lines per family are broadly tested later in development. This data structure leaves breeders with a limited number of lines with the best estimates of phenotypic value, and in addition, these lines often have less direct relatedness to each other, which can make sharing data between individuals challenging. Despite these challenges and limitations, breeding programs need to utilize genomic prediction to accelerate genetic improvement in order to meet future production demands. This project is focused on developing a community resource by utilizing public cooperative regional trial data from MG 00 – IV and genotypic data from available lines to train genomic prediction models. We are interested in analyzing this data using a genomic prediction approach to both help breeders calculate genomic prediction values for selection purposes as well as to help breeders select the best parents to generate populations with favorable population means and variances.

P0750: Legumes, Soybean, Common Bean, and related

Selection for Seed Coat Shininess Elevates Seed Oil Content under Soybean Domestication

Dajian Zhang, Purdue university, West Lafayette, IN, **Shuai Li**, Qingdao Agricultural University, Chengyang District Qingdao, China, Stephen A Swarm, Department of Crop Sciences,, Urbana, IL, Randall Nelson, USDA ARS, Urbana, IL and Jianxin Ma, Purdue University, West Lafayette, IN

Many leguminous species have adapted their seed coat with a layer of powdery bloom that contains hazardous allergens and makes the seeds less visible, offering duel protection against potential predators. Nevertheless, a shiny seed surface without bloom was desirable for human consumption and health, and targeted for selection under domestication. Here we show that seed coat bloom in wild soybeans is mainly controlled by *Bloom1* (*B1*), which encodes a transmembrane transporter-like protein for biosynthesis of the bloom in pod endocarp. The transition from the "bloom" to "no-bloom" phenotypes was achieved through artificial selection of a point mutation that naturally occurred in the coding region of *B1* during soybean domestication. Interestingly, this mutation not only "shined" the seed surface, but also elevated seed oil content in the domesticated soybeans. This study shows pleiotropy as a mechanism underlying the domestication syndrome, and may pave new strategies for development of soybean varieties with increased seed oil content and reduced seed dust.

P0751: Legumes, Soybean, Common Bean, and related

Genetic Analysis of Sucrose Concentration in Soybean using Genotyping-by-Sequencing

Alexandra M. Ficht and Istvan Rajcan, University of Guelph, Guelph, ON, Canada

Soybean (*Glycine max* L. Merr) is an economically important crop, especially in the growing agri-food industry. Understanding the genetic control of specific traits in soybean is required for achieving plant breeding objectives. Sucrose concentration is becoming an increasingly important trait for the production of soy food products. The objectives of this study were: 1) to identify quantitative trait loci (QTL) for sucrose concentration, 2) to explore the allelic variation present in the panel for candidate genes related to sucrose synthesis in soybean, and 3) to determine the effect of genotype and environment on varying sucrose levels. Analysis of molecular data of 266 soybean genotypes grown in three locations, Ridgetown, St. Pauls and Woodstock, ON, Canada, have provided information on the range of sucrose concentration available for the enhancement of food-grade cultivars. QTL analysis has been performed using genotyping-by-sequencing (GBS) data analyzed in both TASSEL and GAPIT. Preliminary results from combined 2015 and 2016 field trials show significant variation in sucrose concentration among

different genotypes in each location-year. A putative QTL has been matched to the genomic region of a known gene involved in sucrose metabolism. Improved understanding of the genes involved in sucrose biosynthesis will allow for its manipulation in future soybean cultivars.

P0752: Legumes, Soybean, Common Bean, and related
Genome-Wide Analysis of Soybean *LBD* genes
Hui Yang, Nanjing Agricultural University, Nanjing, China
Plant-specific *LBD* (*LATERAL ORGAN BOUNDARIES Domain*) genes play critical roles in various plant growth and development processes. However, the number and characteristics of *LBD* genes in soybean [*Glycine max* (L.) Merr.] remain unknown. Here, we identified 90 *LBD* homologous genes in the soybean genome that phylogenetically clustered into two classes. The majority of the *GmLBD* genes were evenly distributed across all 20 soybean chromosomes, and 77 (81.11%) of them were detected in segmental duplicated regions. Furthermore, the exon–intron organization and motif composition for each *GmLBD* were analyzed. A close phylogenetic relationship was identified between the soybean *LBD* genes and 41 previously reported genes of different plants in the same group, providing insights into their putative functions. Expression analysis indicated that more than half of the *LBD* genes were expressed, with the two gene classes showing differential tissue expression characteristics; and they were differentially induced by biotic and abiotic stresses. To further explore the functions of *LBD* genes in soybean, *GmLBD12* was selected for functional characterization. *GmLBD12* was mainly localized to the nucleus and showed high expression in root and seed tissues. Overexpressing *GmLBD12* in *Arabidopsis thaliana* resulted in increases in lateral root (LR) number and plant height. Quantitative real-time polymerase chain reaction (qRT-PCR) analysis demonstrated that *GmLBD12* was induced by drought, salt, cold, indole acetic acid (IAA), abscisic acid (ABA), and salicylic acid (SA) treatments. This study provides the first comprehensive analysis of the soybean *LBD* gene family and a valuable foundation for future functional studies of *GmLBD* genes.

P0753: Legumes, Soybean, Common Bean, and related
A Whole Genome Association Study in Soybean
Fangzhou Zhao, Nanjing Agricultural University, Nanjing, China
Bacterial leaf pustule (BLP) caused by *Xanthomonas axonopodis* pv. *glycines* (Xag) is a worldwide disease. Two resistant genes reported in USA and Korea have been widely utilized. In China, two strains, C5 and B523, were identified from southern China. However, the genetic architecture of soybean lines to the Chinese Xag strain remains to be revealed. To mapping QTL/genes conferring resistance to a highly pathogenic Xag strain C5, we conducted a genome-wide association study by using two soybean natural populations. In a Yangtze-Huai soybean breeding germplasm population with 573 lines, 21 out of 61166 SNPs were detected in association with the disease resistance under a significant level of $-\log_{10} P \geq 4$, and four QTL on chromosome 5 and 17 were extracted. Total 18 significant SNP markers were also identified using an introduction accession population with 271 lines and 35240 SNPs, and six putative QTL distributed on the chromosome 6, 12, 13, 16, 17 and 20 were proposed. Moreover, a locus near the previously reported resistant gene *rxp* on chromosome 17 was detected in both populations. It indicated that *rxp* gene was also the major controller for resistance to the Chinese Xag strain, and more QTL could be involved in the genetic structure of the resistance in soybean. Our results will provide useful information for genetic dissection and molecular breeding of soybean resistance to BLP.

P0754: Legumes, Soybean, Common Bean, and related
Genome-Wide Association Analysis of Cysteine and Methionine Content in Soybean Seeds
Sidiki Malle, Laval University, Quebec, QC, Canada
Soybean is an important source of protein, oils and carbohydrates, as well as other beneficial nutrients. Although they are essential for human nutrition, the seed content in sulfur-containing amino acids (cysteine and methionine) is often poorly characterized. The overall objective of this study was to contribute to the improvement of the nutritional quality of soybeans by gaining a better understanding of the genomic regions controlling these traits. Here, we report the results of a GWAS analysis of cysteine and methionine content in seeds performed on a collection 137 soybean lines that are representative of the diversity of soybeans cultivated in Canada. The phenotypic characterization of the lines was done using NIR and the lines were genotyped using a combined GBS and WGS approach. The software GAPIT was used to identify candidate QTL associated with cysteine, methionine and cysteine + methionine content. GWAS analysis allowed us to detect a total of 10 QTLs distributed on 6 chromosomes (Chr5, Chr8, Chr10, Chr14, Chr19 and Chr20), each of which explained from 14 to 23% of the total phenotypic variation. In some cases, the same QTL were detected for all three traits (Chr10 and Chr20) and accounted for more than 35% for the phenotypic variation. These results provide insight into the genetic basis for the accumulation of these sulfur-containing amino acids in soybeans. QTLs identified in this study can be used for marker-assisted selection to improve the nutritional quality of soybeans and help in the identification of candidate genes.

P0755: Legumes, Soybean, Common Bean, and related
Transcriptional Profiling of Mechanically and Genetically Sink-Limited Soybean
Anne V. Brown, USDA-ARS-CICGRU, Ames, IA and Karen A Hudson, USDA-ARS, West Lafayette, IN
The absence of a reproductive sink causes physiological and morphological changes in soybean plants. These include increased accumulation of nitrogen and starch in the leaves and delayed leaf senescence. To identify transcriptional changes that occur in leaves of these sink-limited plants, we used RNAseq to compare gene expression levels in trifoliate leaves from depodded and ms6 male sterile soybean plants and control plants. In both sink-limited tissues, we observed a deferral of the expression of senescence-associated genes and a continued expression of genes associated with leaf maturity. GO-terms associated with growth and development and storage proteins were over-represented in genes that were differentially expressed in sink-limited tissues. We also identified bHLH, ARF, and SBP transcription factors expressed in sink-limited tissues, while the senescing control leaves expressed WRKY and NAC transcription factors. We identified genes that were not expressed during normal leaf development but that were highly expressed in sink-limited plants, including the SGR3b “non-yellowing” gene. These differences highlighted several metabolic pathways that were involved in distinct modes of resource partitioning of leaves with the “stay green” phenotype.

P0756: Legumes, Soybean, Common Bean, and related

Characterizing and Comparing International Germplasm Collections: A Database of Soybean Variant Data.

Anne V. Brown, Shawn Conners, Nathan Weeks, David Grant and Steven B. Cannon, USDA-ARS-CICGRU, Ames, IA

There are over 230,000 soybean accessions in germplasm repositories worldwide, making the identification of truly unique accessions difficult. High throughput genotyping costs have dropped sufficiently to enable dense genotyping of large germplasm collections. Nevertheless, large challenges remain due to the sheer volume of such genotype data. Comparisons between genotyping projects are additionally complicated by lack of common markers among data sets, differences in accession names, SNPs called from different reference genomes, and by inconsistent data formats. Here we describe a new database for genotyping data as a partial replacement for dbSNP for soybean. Using genotyping datasets from published U.S., Brazilian, Canadian, Chinese, and Korean soybean, we identified common loci and accessions between the datasets. Among the SNP arrays, we found that overlap of SNP positions is as high as 85% between the arrays. All SNPs are assigned a new SNP ID and SNPs common between multiple datasets will have the same identifier. Old names (ex. rs and ss IDs) are kept and will be searchable. This database will be the foundation for developing new interactive tools for soybean breeders at SoyBase.

P0757: Legumes, Soybean, Common Bean, and related

Genetic Association Analysis and Genomic Prediction of Soluble Carbohydrates in Soybeans

Tri D. Vuong[1], Gunvant Patil[1,2], Babu Valliyodan[1], Haiying Shi[1], J. Grover Shannon[3], Pengyin Chen[3] and Henry T. Nguyen[1], (1)University of Missouri, Columbia, MO, (2)University of Minnesota, St Paul, MN, (3)University of Missouri, Portageville, MO

Soybean [*Glycine max* (L.) Merr.] seed contains approximate 35% of soluble sugar components, including sucrose, raffinose, and stachyose. Among these, sucrose is easy to digest and a desirable trait for taste and flavor. In contrast, raffinose and stachyose are difficult to digest by monogastric animals and act as anti-nutritional factors; thus, reducing raffinose and stachyose biosynthesis is considered as a key quality trait goal in soy food and feed industries. Soybean germplasm with low stachyose content was identified and employed in soybean breeding; however, more efforts need to be made to improve carbohydrate components. The objective of this study was to discover new sources of carbohydrates with desirable combinations of these three soluble sugar components in an effort to identify and characterize genomic regions or genes controlling these traits. The evaluations of exotic soybean germplasm identified many plant introductions with elevated sucrose (>8.0%) and reduced raffinose and stachyose (<2.5%). Linkage genetic analysis conducted in a bi-parental mapping population detected and mapped major quantitative trait loci for high sucrose and low stachyose. Genome-wide association mapping of a diverse germplasm panel using a whole-genome sequence-based DNA marker data set identified several single nucleotide polymorphism (SNP) loci significantly associated with high sucrose and low stachyose phenotypes in new sources. Genomic prediction using the diverse germplasm panel as a training population was subsequently performed to identify new soybean germplasm with desirable sugar components. The significant SNP loci and new accessions are valuable resources, facilitating the improvement of soluble carbohydrates in soybean seeds.

P0758: Legumes, Soybean, Common Bean, and related

Next-Generation Sequencing from Bulked-Segregant Analysis Accelerates the Simultaneous Identification of Two Qualitative Genes in Soybean

Yong Guo, Jian Song and Lijuan Qiu, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China

Next-generation sequencing (NGS)-based bulked-segregant analysis (BSA) approaches have been proven successful for rapidly mapping genes in plant species. However, most such methods are based on mutants and usually only one gene controlling the mutant phenotype is identified. In this study, NGS-based BSA was employed to map simultaneously two qualitative genes controlling cotyledon color of seed in soybean. Yellow-cotyledon (YC) and green-cotyledon (GC) bulks from progenies of a biparental population (Zhonghuang30 ´ Jiyu102) were sequenced. The SNP-index of each SNP locus in YC and GC bulks was calculated and Δ(SNP-index) was used to identify two genomic regions on chromosomes 1 and 11 harboring respectively loci *qCC1* and *qCC2*. These two BSA-seq-derived loci were further validated with SSR markers and fine-mapped. *qCC1* was mapped to a 30.7-kb region containing four annotated genes and *qCC2* was mapped to a 67.7-kb region with nine genes. These two regions contained respectively genes *D1* and *D2*, which had previously been identified by homology-based cloning as being associated with cotyledon color. Sequence analysis of the NGS data also identified a frameshift deletion in the coding region of *D1*. These results suggested that BSA-seq could accelerate the mapping of loci controlling qualitative traits, even if a trait is controlled by more than one locus.

P0759: Legumes, Soybean, Common Bean, and related

Cloning and Functional Diversification of Genes underlying Photoperiodic Response and Maturity in Soybean and other Related Species in Legume

Zhengjun Xia[1], Xinzheng Zhang[2], Hong Zhai[2], Hongyan Wu[2], Jiajia Zheng[2], Yaying Wang[2], Bo Hu[2], Yuqiu Li[2] and Yanhua Li[2], (1)Northeast Institute of Geography and Agroecology, CAS, Harbin, China, (2)Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China

Gene regulatory networks involved in flowering time and photoperiodic responses in legumes remain unknown. Although the major maturity gene *E1* has been successfully deciphered in soybean, knowledge on the functional conservation of this gene is limited to a certain extent to *E1* homologs in legumes. The ectopic expression of *Phvul.009G204600* (*PvE1L*), an E1 homolog from common bean, delayed the onset of flowering in soybean. By contrast, the ectopic expression of *Medtr2g058520* (*MtE1L*) from *Medicago truncatula* did not affect the flowering of soybean. Characterization of the late flowering *mte1l* mutant indicated that *MtE1L* promoted flowering in *Medicago truncatula* . Moreover, all transgenic *E1*, *PvE1L* and MtE1L soybean lines exhibited phenotypic changes in terms of plant height. Transgenic *E1* or *PvE1L* plants were taller than the wild-type, whereas transgenic *MtE1L* plants produced dwarf phenotype with few nodes and short internodes. Thus, functional conservation and diversification of E1 family genes from legumes in the regulation of flowering and plant growth may be associated with lineage specification and genomic duplication.

P0760: Legumes, Soybean, Common Bean, and related

Exploring Soybean Germplasm for Drought and Agronomic-Related Traits Based on the True Breeding Values

Avjinder Kaler and Larry C. Purcell, University of Arkansas, Fayetteville, AR

Selection of parents is one of the main decisions faced by plant breeders. True breeding values (TBVs) can be a guide in selecting parental genotypes in a breeding programs. True breeding values were calculated using the results from genome wide association studies (GWAS) of drought related traits including canopy temperature (CT), canopy wilting (CW), carbon isotope ratio ($\delta^{13}C$), oxygen isotope ratio ($\delta^{18}O$), and nitrogen derived from atmosphere (NDFA), measured from our research along with several traits from GRIN including seed shattering (SS), height (Ht), lodging (LG), oil (OL), protein (PR), seed weight (SW), stem termination (ST), and yield (Y). Accessions (373) used for GWAS were considered a training set and the rest of the accessions in the germplasm collection were considered a validation set. Over 14,000 accessions from the germplasm collection have phenotypic data for agronomic traits. From phenotyped accessions, two traits, ST and SW, which are not affected greatly by environmental variation and sampling error, were used to check the accuracy by correlating actual phenotype and TBVs. We observed highly significant positive correlations between actual phenotype and TBVs for ST (r = 0.60) and SW (r = 0.56). Based on TBVs, we found that the expected correlations between agronomic traits such as yield were significantly correlated with height (r = 0.50) and oil (r = 0.56), and negatively correlated with protein (r = -0.23), early seed shattering (r = -0.68), late seed shattering (r = -0.48), and seed weight (r = -0.37). Similarly, we observed the expected correlations between drought-related traits including a positive correlation between CW and CT (r = 0.60), and a positive correlation of $\delta^{13}C$ with CW (r = 0.53) and CT (r = 0.67). Extreme drought-tolerant accessions were identified, which can be used as potential parents in a breeding program.

P0761: Legumes, Soybean, Common Bean, and related

Development of High Value Soybeans with High Oleic Acid Vegetable Oil and Enhanced Nutritional Energy Meal

Jeong-Hwa Kim, University of Missouri-Columbia, Columbia, MO and Kristin Bilyeu, USDA-ARS, University of Missouri, Columbia, MO

Soybean (*Glycine max* L. Merrill) is one of the important crops worldwide which has been widely used as an edible source for humans and animals because of its high nutritional value. Soybean grain generally contains 40% protein, 20% oil, 35% carbohydrates, and other useful secondary metabolites. Among these components, the fatty acids in the seed oil and carbohydrate compounds in the seed meal deserve attention when considering soybean value. The fatty acids in soybean oil are generally comprised of 11% palmitic, 4% stearic, 23% oleic, 54% linoleic, and 8% linolenic acid. For oil, the high oleic and low linolenic acid traits (HOLL; >70% oleic and <3% linolenic acid) are now targeted to improve oxidative stability of soybean oil and recapture lost market value due to issues with *trans* fats. In terms the carbohydrate compounds of soybean meal, there are three major oligosaccharides; sucrose which is fully digested by monogastric animals and related to sweetness of soy-based foods, and raffinose and stachyose (RFOs) which cannot be digested in the animals. Research related to the carbohydrates has been conducted to elevate sucrose, and to reduce raffinose and stachyose (low RFO trait). The objective of this research is to develop soybean germplasm that can increase the value of soybean oil and meal. Soybean lines are being developed with the aid of marker assisted selection for combinations of genes that contain the four alleles necessary for the HOLL trait, one allele for the low RFO trait, and the appropriate targeted maturity group (III and IV). Soybean germplasm with the desired allele combinations will be grown in the targeted environment to produce seed that can be assessed for increased value in the oil and meal. The results generated will determine the feasibility of soybean variety development with this unique combination of oil and meal traits.

P0762: Legumes, Soybean, Common Bean, and related

Detection of QTL Underlying Seed Quality Components in Soybean [*Glycine max* (L.) Merr.]

Masum Akond[1], Jiazheng (John) Yuan[1], Shiming Liu[2], Stella Kantartzi[3], Khalid Meksem[4], Nacer Bellaloui[5], David A. Lightfoot[3] and **Abdelmajid Kassem**[1], (1)Fayetteville State University, Fayetteville, NC, (2)Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL, (3)Southern Illinois University, Carbondale, IL, (4)Southern Illinois University Carbondale, Carbondale, IL, (5)USDA-ARS, Stoneville, MS

Improving seed composition and quality, including protein, oil, fatty acids, and amino acids content is an important goal of soybean farmers and breeders. The aim of this study was to map the QTL for protein, oil, fatty acids, and amino acids content with 5,376 single nucleotide polymorphism (SNP) markers using the 'Hamilton' by 'Spencer' recombinant inbred line (RIL) population (H S, n = 93). A total of 13 QTLs for the traits studied have mapped on 3 chromosomes (Chr.) of the soybean genome. Three major QTLs have been mapped to a 7–13 cM region on Chr 6. One major QTL for oil content (qOIL001) that explained approximately 76% of the total variation in oil content in this population defined by SNP; one major QTL for amino acid Alanine (ALA) (qALA001) that explained approximately 74% of the total variation in ALA content in this population; moreover, two major QTL for palmitic acid (qPAL001 and qPAL002) were identified on Chr. 6 and explained approximately 21% of palmitic acid content in this population. SNP markers closely linked to the QTLs and QTLs identified in this study will be a useful tool to soybean breeders to develop and select soybean lines with higher seed composition qualities, using marker-assisted selection.

P0763: Legumes, Soybean, Common Bean, and related

Stability Analysis of Soybean Seed Protein Concentration in the Quality Traits Tests

Siddhi J. Bhusal, University of Minnesota, St Paul, MN, George L. Graef, Dept. of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, Charles Hurburgh, Dept. of Agric. and Biosystems Engineering, Iowa State University, Ames, IA and Aaron Lorenz, University of Minnesota, St. Paul, MN

Soybean seed is an important source of protein for human and animal consumption. Extensive research has been accomplished on the measurement and genetics of soybean seed protein concentration, which has helped facilitate development of varieties with higher seed protein concentration. Very little work, however, has been done on studying the stability of soybean seed protein concentration across environments. To help fill this gap, we performed a stability analysis of soybean seed protein concentration using data collected as part of the Regional Quality Traits Test. The data from these trials is ideal for this analysis because of the wide range of genetic variation for protein concentration included, as well as the large number of maturity group zones, locations, and years represented. Analysis of seed protein concentration of soybean genotypes from nine maturity groups (MG) tested in various locations in the Unites States and Canada from 2002 to 2016 revealed

extensive genotype-by-environment interactions. A Finlay-Wilkinson regression analysis indicated that approximately 60% of the genotypes analyzed displayed significant responses to environment for protein concentration (i.e., $\beta_1 > 0$, $p<0.05$). Tests of heterogeneity of regression coefficients among genotypes within trials were significant for about two third of the cases. This result indicates that genetic variation for protein concentration stability is prevalent. The correlations between average seed protein concentration and β_1 (measure of Type I and II stability) was not significantly different for at least 75% of the trials, and the sign of the correlation coefficient was not consistent in the remaining trials. The correlations between average seed protein concentration and the regressions R^2 (measure of Type III stability) also were not significantly different from 0 in at least 75% of the trials. These results suggest that variation in average seed protein concentration among soybean varieties is not related to variation in protein stability among varieties.
Acknowledgements: We are indebted to all the cooperators who have grown these trials through the years, and the United Soybean Board for financial support of the Quality Traits tests. These data and findings would not be possible without the cooperation of soybean breeding programs throughout North America.

P0764: Legumes, Soybean, Common Bean, and related
Whole-Genome Analysis of Parental Stocks and Random Lines Derived from *G. max* × *G. soja* Crosses Provides a Reference to Identify Genomic Regions Controlling Traits Associated with Domestication
Qijian Song[1], Charles Quigley[1], Edward Fickus[1], Earl Taliercio[2], Thomas Carter[2], Andrew Scaboo[3], Steven G. Schroeder[4] and Márcio E. Ferreira[5], (1)USDA-ARS, Soybean Genomics and Improvement Lab, Beltsville, MD, (2)Soybean and Nitrogen Fixation Research, USDA-ARS, Raleigh, NC, (3)University of Missouri, Columbia, MO, (4)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (5)5Embrapa Labex-USA, Beltsville, MD
Pairwise genetic distances among the 1,168 *G. soja* accessions were calculated based on the SNPs in the SoySNP50K Illumina Beadchip and were used to classify the accessions into 81 clusters. We selected one accession with the largest average distance to other accessions from each cluster and assembled a wild soybean core collection which contains >90% of the diversity of the entire USDA-ARS wild soybean collections. A number of wild soybean accessions from the core collection were crossed with cultivated soybean. We sequenced the accessions in the core collections, parental stock and random lines of the *G. max* × *G. soja* crosses created at USDA-ARS, NC, and University of Missouri, and identified SNPs among wild accessions, between wild soybean and cultivated soybean parents, and among the random lines. Further analyses discovered SNPs with fixed or near fixed alleles in the *G. soja* vs. *G. max* parents and *G. soja* vs. random lines. This study provides a baseline for the discovery of the genomic regions controlling desirable traits affected by wild soybean and traits under domestication.

P0765: Legumes, Soybean, Common Bean, and related
Genome-Wide Comparative Analysis of DNA Methylation between Soybean Cytoplasmic Male-Sterile Line NJCMS5A and Its Maintainer NJCMS5B
Yanwei Li, Xianlong Ding, Xuan Wang, Tingting He, Hao Zhang, Longshu Yang, Tanliu Wang, Linfeng Chen, Junyi Gai and **Shouping Yang**, Soybean Research Institute, National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean (Ministry of Agriculture), State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China
DNA methylation is an important epigenetic modification. It can regulate the expression of many key genes without changing the primary structure of the genomic DNA, and plays a vital role in the growth and development of the organism. The genome-wide DNA methylation profile of the cytoplasmic male sterile (CMS) line in soybean has not been reported so far. In this study, genome-wide comparative analysis of DNA methylation between soybean CMS line NJCMS5A and its maintainer NJCMS5B was conducted by whole-genome bisulfite sequencing. The results showed 3527 differentially methylated regions (DMRs) and 485 differentially methylated genes (DMGs), including 353 high-credible methylated genes, 56 methylated genes coding unknown protein and 76 novel methylated genes with no known function were identified. Among them, 25 DMRs were further validated that the genome-wide DNA methylation data were reliable through bisulfite treatment, and 9 DMRs were confirmed the relationship between DNA methylation and gene expression by qRT-PCR. Finally, 8 key DMGs possibly associated with soybean CMS were identified.

P0766: Legumes, Soybean, Common Bean, and related
Identifying Genes Important for Determining Lateral Branch Angle in Soybean
Suma Sreekanta[1], Anna N. Hofstad[1], Kamaldeep S. Virdi[1], Fengli Fu[1], Jean-Michel Michno[1], Adrian O. Stec[1], Jeffrey Roessler[1], Benjamin W. Campbell[1], Aaron Lorenz[2], Robert M. Stupar[1] and Gary J. Muehlbauer[2], (1)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (2)University of Minnesota, St. Paul, MN
Canopy architecture of plants is an important determinant of light interception and consequently photosynthesis, and ultimately yield in crop species. We have explored the variation in canopy architecture in a fast neutron (FN) mutagenized population of soybean as well as in a set of diverse lines from the USDA Soybean Germplasm Collection. We have observed variation in both populations for several traits, such as lateral branch angle, branching density, branch orientation and overall plant shape. Changes in shoot architecture and particularly those that display differences in the angle of lateral branch are of particular interest. In the FN collection, we have identified mutants that were altered in branch angle. Our aim is to identify, map and clone the loci responsible for determining branch angle in soybeans. We are using a complementary approach of bulk segregant whole genome sequencing and array Comparative Genomics Hybridization to map regions of the genome controlling the phenotype. Using this approach, we have identified candidate genes in the mapped regions that may be important for controlling lateral branch angle. Experiments are underway to functionally characterize the candidate genes as well as to determine the effect of lateral branch angle on overall yield of soybean. Additionally, association mapping with the USDA Soybean Germplasm Collection has identified regions with significant associations with the branch angle phenotype.

P0767: Legumes, Soybean, Common Bean, and related
Genetic Diversity Patterns and Domestication Origin of Soybeans

Soon-Chun Jeong, Korea Research Institute of Bioscience and Biotechnology, Chungbuk, Korea, Republic of (South) and Jung-Kyung Moon, 2National Institute of Crop Science, Rural Development Administration, Jeonju-si, Korea, Republic of (South)
Understanding diversity and evolution of a crop is an essential step to implement a strategy to expand its germplasm base for crop improvement research. Samples intensively collected from Korea, which is a small but central region in the distribution geography of soybeans, were genotyped to provide sufficient data to underpin genome-wide population genetic questions. After removing natural hybrids and duplicated or redundant accessions, we obtained a non-redundant set comprising 1,957 domesticated and 1,079 wild accessions to perform population structure analyses. Genetic diversity of Korean soybeans appeared to be as high as those of Chinese or Japanese soybeans. Correlations between genetic distance and geographic distance were strong in wild soybean populations but weak in domesticated soybean populations. Our analysis demonstrates that while wild soybean germplasm will require additional sampling from diverse indigenous areas to expand the germplasm base, the current domesticated soybean germplasm is saturated in terms of genetic diversity. We then showed that our genome-wide polymorphism map enabled us to detect genetic loci underling flower color, seed-coat color, and domestication syndrome. A representative soybean set consisting of 194 accessions was used to infer phylogenetic relationships among soybean subgroups without sampling bias. The soybean accessions were divided into one domesticated subpopulation and four wild subpopulations that could be traced back to their geographic collection areas. Population genomics analyses suggested that the monophyletic group of domesticated soybeans was originated in eastern Japan. The results were further substantiated by a phylogenetic tree constructed from domestication-associated single nucleotide polymorphisms identified in this study.

P0768: Legumes, Soybean, Common Bean, and related
Genetic Diversity Analysis of South African Soybean Genotypes using Agronomic and Nutritional Quality Traits
Maletsema Alina Mofokeng, Agricultural Research Council-Grain Crops, Potchefstroom, South Africa
Soybean is one of the most important leguminous crops grown for human and animal consumption as well as for development of industrial products. It is essential for protein, oil, mineral content and vitamins. Knowledge of genetic diversity among soybean genotypes is essential for current and future breeding. The objective of this study was to assess the level of genetic diversity present among South African soybean genotypes using agro-morphological and nutritional quality traits. Ninety-six soybean genotypes were planted in Potchefstroom during the 2016/17 soybean-growing season in an alpha lattice replicated twice. Data on quantitative and nutritional quality traits were subjected to principal component (PC), hierarchal cluster, and multivariate analyses and a dendrogram was constructed using the Unweighted Pair Group Method with Arithmetic Mean. The principal component analysis revealed three most important PCs contributing 63.19%, 25.43% and 8.88% to the total variation of 97.5%, respectively. The hierarchical clustering revealed three major clusters with further sub-clusters. The accessions 2015/06/12, 69 S 10, PR 154-14, R 5-4-2 M, Hawkeye (USSR), and PR 145-2 were the most diverse. There were significant differences among the accessions based on nutritional quality traits such as oil, protein and stearic acid across the locations. The protein content varied from 29.1% to 35.6%, oil content varied from 10.6% to 20.7% whereas oleic acid and ash varied between 6.8% and 30.8%, and 4.3% and 8.2%, respectively. There was vast genetic diversity among the soybean genotypes. The presence of genetic diversity will aid breeders in selections and hybridization programmes for crop improvement.
Keywords: Agro-morphology, genetic diversity, nutritional quality, soybean

P0769: Legumes, Soybean, Common Bean, and related
Identification of Quantitative Trait Loci Associated with Seed Protein Concentration in High-Protein Soybean Recombinant Inbred Line Populations
Rachel Whaley[1], Milad Eskandari[1] and Lewis Lukens[2], (1)University of Guelph-Ridgetown, Ridgetown, ON, Canada, (2)University of Guelph, Guelph, ON, Canada
Soybean (*Glycine max* (L.) Merrill) is a predominant global source of plant-based dietary protein for human and livestock consumption. Western society has increasingly embraced soy as a more healthful substitute for meat and dairy products. Manipulating the quantity and quality of protein in soybean seeds can alter the nutritional value of soy-food products. Protein concentration is a complex trait that is negatively associated with yield, which discourages the production of high-protein soybean cultivars through classical phenotypic selections. Quantitative trait loci (QTL) can be used to expedite the improvement of protein content, while mitigating yield loss. The objective of this study is to identify QTL associated with seed protein concentration in recombinant inbred line (RIL) populations derived from high-protein parental cultivars. Two RIL populations were created from crosses between AC X790P (49% protein concentration, dry basis), and the elite Canadian cultivars, S18-R6 (40% protein concentration, dry basis) and S23-T5 (41% protein concentration, dry basis). Approximately 190 RILs were used for QTL mapping in each population. The RILs were evaluated for yield, seed protein concentration, and protein meal concentration at three locations for two years, comprising five environments. The populations were genotyped using genotyping-by-sequencing to identify single nucleotide polymorphism (SNP) markers. Numerous protein-related QTL (>10% explained variation) were identified, and cross-validated between the RIL populations. The identification of QTL associated with elevated seed protein concentration would be of immense value to the development of new high-yielding soybean cultivars with improved nutritional value using marker-assisted selection.

P0770: Legumes, Soybean, Common Bean, and related
Genetic Mapping and Validation of the Loci Controlling 7S α' and 11S A-Type Storage Protein Subunits in Soybean [*Glycine max* (L.) Merr.]
Jeffrey D Boehm Jr[1], Vi Nguyen[2], Rebecca Tashiro[3], Dale Anderson[4], Chun Shi[4], Xiaoguang Wu[4], Lorna Woodrow[4], Kangfu Yu[4], Yuhai Cui[2] and Zenglu Li[5], (1)Institute of Plant Breeding, Genetics & Genomics, The University of Georgia, Athens, GA, (2)Agriculture and Agri-Food Canada, London, ON, Canada, (3)Institution of Plant Breeding, Genetics & Genomics, The University of Georgia, Athens, GA, (4)Agriculture and Agri-Food Canada, Harrow, ON, Canada, (5)Institute of Plant Breeding, Genetics & Genomics and Department of Crop and Soil Sciences, The University of Georgia, Athens, GA
The storage protein globulins β-conglycinin (7S subunit) and glycinin (11S subunits) can affect the quantity and quality of proteins found in soybean seeds and account for more than 70% of total soybean protein. Manipulating the storage protein subunits to enhance soymeal nutrition

and for desirable tofu manufacturing characteristics are two end-use quality goals in soybean breeding programs. To aid in developing soybean cultivars with desired seed composition, an F_2 mapping population (n = 448) and an F_5 RIL population (n = 180) were developed by crossing high protein cultivar 'Harovinton' with the breeding line SQ97-0263_3-1a, which lacks the 7S α', 11S A_1, 11S A_2, 11S A_3 and 11S A_4 subunits. Storage protein composition of each individual in the F_2 and F_5 populations were profiled using SDS-PAGE. Based on the presence/absence of the subunits, genomic DNA bulks were formed among the F_2 plants to identify genomic regions controlling the 7S α' and 11S protein subunits. By utilizing polymorphic SNPs between the bulks characterized with Illumina SoySNP50K iSelect BeadChips at targeted genomic regions, KASP assays were designed and used to map QTLs causing the loss of the subunits. Soybean storage protein QTLs were identified on Chromosome 3 (11S A_1), Chromosome 10 (7S α' and 11S A_4), and Chromosome 13 (11S A_3), which were also validated in the F_5 RIL population. The results of this research could allow for the deployment of marker-assisted selection for desired storage protein subunits by screening breeding populations using the SNPs linked with the subunits of interest.

P0771: Legumes, Soybean, Common Bean, and related
Identification and Characterization of Fast Neutron Mutant Soybean Lines with Altered Seed Composition for Improvement of Seed Composition
Elizabeth Prenger, Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Athens, GA, Rouf Mian, USDA-ARS, NC State University, Raleigh, NC, Robert M. Stupar, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN and Zenglu Li, Institute of Plant Breeding, Genetics & Genomics and Department of Crop and Soil Sciences, The University of Georgia, Athens, GA
Improving seed composition is a key breeding goal in modern soybean breeding. Protein and oil concentration, amino acid profile, fatty acid content, and carbohydrate composition play a vital role in the utility of soybean for food, feed, and fuel. A meal protein content of 48% or higher is desired for animal feed without sacrificing oil content. These criteria can be difficult for breeders to meet due to a negative relationship between protein and oil contents, and between protein content and yield. Ideal soybean seed also contains high concentrations of essential amino acids, a healthier fatty acid profile, and high levels of digestible sucrose. Fast neutron (FN) irradiation induces genomic deletions, duplications, and translocations in soybean resulting in mutant phenotypes. In this study, FN irradiation was used to develop two soybean mutant populations that were screened for altered seed composition phenotypes. Results from a total of five environments in two years and wet chemistry validation identified 40 stable mutant phenotypes for protein, oil, and sucrose contents. Twenty-three mutant lines have been entered into 2017 yield trials to determine the impact of genomic changes on seed yield. Comparative genomic hybridizations (CGH) of four mutants were performed to identify putative genomic regions responsible for the mutations, and whole genome sequencing will provide additional information. Bi-parental populations have been created to confirm these genomic regions. Identification of underlying genomic changes in seed composition mutants and generation of new mutant breeding materials could enhance breeding efforts for improvement of seed composition in new soybean varieties.

P0772: Legumes, Soybean, Common Bean, and related
Harnessing and Utilization of Untapped Genetic Diversity in Soybean
Yinghui Li, Chinese Acdemy of Agricultural Sciences, Beijing, China
As the progenitor of soybean, wild soybean (*Glycine soja* Sieb. and Zucc.) is a promising source of novel genes for soybean breeding that has been mostly untapped. To identify diversity in wild soybean, a serial of attempts have been made using re-sequencing approach and a series of genomic region and genes with selection signals during domestication had been revealed. To identify a wide range of nucleotide and structural variations conserved in wild soybean, we established and analyzed a pan-genome of G. soja by sequencing and de novo assembly of seven phylogenetically and geographically representative accessions. The assembly-based pan-genome enabled us to identify more variations than by re-sequencing alone. This data was very useful to discover candidate genes for a variety of traits of agronomic value including resistance to soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe), a highly destructive pathogen of soybean. A genotyping chip with 768 SNPs was used to genotype both a panel of diverse accessions and a set of RILs. Combining linkage, association mapping and join linkage-association mapping, a new minor resistant locus, *SCN3-11* was cross-validated. *SCN3-11* was the closest paralog of major resistant *rhg1-b* cloned. Our research showed that the combination of rhg1-b and *SCN3-11* provides a high level of SCN resistance.

P0773: Legumes, Soybean, Common Bean, and related
Adaptation of Soybean in Northern Ghana for Smallholder Farmers
Carrie Miranda, University of Missouri, Columbia, MO and Kristin Bilyeu, USDA-ARS, University of Missouri, Columbia, MO
There is a high demand for soybean in African countries, but available varieties are poor yielding. This can be partially attributed to inadequate adaptation of soybean to a tropical climate. Adaptation will require knowledge of allelic combinations of the characterized maturity genes: E1, E2, E3, and E4; the long juvenile trait, and stem architecture. The long juvenile trait influences flowering time in short days, which characterize low latitudes. Stem architecture includes the determinate or indeterminate phenotypes. By understanding the influence of these genetic components on adaptation, it may be possible to control season length and improve yield greater than the currently available African varieties. To achieve this objective, six populations were initiated in which the genes of interest were segregating. 260 recombinant inbred lines were created across the six populations and were field tested in 5 locations in northern Ghana in 2016. During this time phenotypes for flowering, maturity, and other agronomic traits were noted. Our initial results from one population suggest the long juvenile trait plays the most influential role on days to flower over E1. However, across populations segregating for the long juvenile trait these data also insinuate that that different alleles of this gene may also influence flowering phenotypes. Further analysis is being conducted to understand the effect of maturity gene allelic combinations on season length. The combined knowledge of the genetic control of these traits will allow local Ghanaian breeders to produce varieties that can cater to the needs of small farmers in the north.

P0774: Legumes, Soybean, Common Bean, and related
Transcriptome and Epigenomic Regulation in Salt-Sensitive (Union) and Salt-Tolerant (Lee-68) Genotypes of Soybean

LaMont A. Croom, Dyneesha C. Head, Zachary B. Gossett, Govind C. Sharma and Venkateswara R. Sripathi, Alabama A&M University, Normal, AL
Soybean (*Glycine max* L.) is a commercially important crop that supports food, feed, and fuel industries. It has a genome size of 1,115 Mb. The United States produced over 117 million metric tons of soybeans in 2016-17 growing season, making it the world's leading agricultural producer of soy. Salinity stress impedes growth, development, and inhibits vegetative growth. About 10 million hectares annually are lost due to salinization and waterlogging. In the United States, approximately 30% of all irrigated land is affected by salinity, compared to the 50% of the farmland area affected by salinity worldwide. DNA methylation affects gene expression and is partly responsible for the phenotypic variation among different crop species. This experiment included two genotypes (Union and Lee-68), three treatment conditions (0, 75 mM and 150 mM NaCl), three collection time points (0 DAT, 7 DAT and 14 DAT) and three replicates (R1, R2, and R3) to generate 54 RNA-Seq and BS-Seq libraries each that were sequenced on Illumina Hiseq2500 and generated over 800 million paired-end reads (2 x 100 bp). Our preliminary analyses in bean revealed that the % cytosine methylation of CG, CHG and CHH contexts ranged between 55-60%, CHG 25-30%, and CHH 5-10%. The methylation patterns in the promoter, genic and repeat regions identified were similar to other plants species. We identified higher CG methylation levels than CHG and CHH contexts. Differentially methylated regions (DMRs) for CG, CHG, and CHH were assessed between the two genotypes. This is the first genome-wide DNA methylation profiling study in salt-sensitive (Union) and salt-tolerant (Lee-68) genotypes of soybean using NGS approach. Furthermore, the comparative unstressed and salt-stressed methylomes and gene expression analyses of Lee-68 and Union soybean genotypes will aid in identifying critical genes and pathways involved in ion flux, ROS (reactive oxygen species) and jasmonate signaling pathways.

P0775: Legumes, Soybean, Common Bean, and related
Effects of an *EPSPS*-Transgenic Soybean Line ZUTS31 on Root-Associated Bacterial Communities during the Field Growth
Yonghua Yang[1], Guihua Lu[1], Chengyi Tang[1] and Xiaomei Hua[2], (1)Nanjing University, Nanjing, China, (2)Nanjing Institute of Environmental Sciences, MEP, Nanjing, China
The worldwide commercial cultivation of transgenic crops, including transgenic soybeans with glyphosate-tolerance, has increased widely during the past 20 years. This increased cultivation is accompanied with potential effects on the environment, including the soil microbial communities, because many rhizosphere and endosphere bacteria play important roles in promoting plant health and growth. Previous studies found that transgenic plants exert differential effects on soil microbial communities, especially rhizobacteria. Additionally, recent studies have comprehensively investigated the root-associated microbiota of model plants, including Arabidopsis thaliana and Lotus japonicas. However, few studies have deeply analyzed the root-associated (endosphere, rhizoplane, and rhizosphere) microbiota of crops in field growth conditions using next generation sequencing. In the present study, soybean root-associated bacterial communities was compared between a 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)-transgenic soybean line ZUTS31 (or simply Z31) and its recipient cultivar Huachun3 (or simply HC3) at the vegetative, flowering, and seed-filling stages. High-throughput sequencing of 16S rRNA gene V4 region amplicons via Illumina MiSeq was used in the analysis. Our results indicated that the EPSPS-transgenic soybean line Z31 exerted no significant effect on the overall alpha and beta diversity of the rhizosphere and root interior (endosphere) plus root surface (rhizoplane) bacterial communities compared with its recipient HC3 at the three developmental stages during a single field growth. Furthermore, soybean development evidently changed the root-associated main nitrogen-fixing bacterial genera, especially from the flowering stage to the seed-filling stage.

P0776: Legumes, Soybean, Common Bean, and related
An Allelic Variation Controlling a High Ratio of Four Seeds *per* Pod Was Utilized to Breed the Excellent Soybean Lines with High Yield
Baoge Zhu, Zhixi Tian, Guoan Zhou, Yi Pan and Xiuwen Chen, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
The number of seeds per pod is one of the critical components related to soybean yield. Increasing the number of seeds per pod without reducing seed size is of great significance for enhancing soybean yield potential. In this study, we utilized an allelic variation (a base G mutated to C in Gmln gene) controlling a high ratio of four seeds per pod in total pods per plant previously identified by us(JGG, 2013) for further enhancing yield potential of soybean cultivars. On the basis of genome sequencing, we used suitable soybean donors (Kefeng14 or Heinong40) with a high ratio of four seeds per pod and containing the allelic variation above to be hybridized with the widely released soybean variety(Kedou1 or Zhonghuang13) without this allelic variation in Huang-Huai-Hai region of China, and then screened the target progenies to be backcrossed with a receptor parent in three generations and further selfing breeding, and obtained a number of elite lines with target phenotype of a high ratio of four seeds per pod and the genotype containing the allelic variation. Finally we developed several excellent lines including KH5-2□KZ5-1, KZ5-5, LZ904-1, and LZ904-7 with a high ratio of four seeds per pod and high yield. Among them, KH5-2 and LZ904-7 exhibited the highest yields in all lines. The field experiment results indicated that LZ904-7 and KH5-2 showed high yields of 3093.0 kg/ha and 3225.0 kg/ha in display area of 3.3 ha, increasing 8.24% and 10.43% compared to that of control variety, respectively.

P0777: Legumes, Soybean, Common Bean, and related
Omics-Based Analysis of Soybean Seedlings Under the Controlled Environment
Kai Uchida[1], Yuji Sawada[1], Koji Ochiai[2], Muneo Sato[1], Mami Okamoto[1], Yutaka Yamada[1], Akane Sakata[1] and Masami Yokota Hirai[3], (1)RIKEN Center for Sustainable Resource Science, Yokohama, Japan, (2)Daiz Energy Co., Ltd., Kumamoto, Japan, (3)RIKEN CSRS, Yokohama, Japan
Plant specific metabolites, which are known to be induced by various environmental stresses, are essential for the adaptation to environments. Such stress-inducible metabolites attract attention as environmental marker for plant breeding, and the various structure-related bioactivities have potential to contribute to natural-product-based drug discovery. Toward stable production of such stress-inducible metabolites, we established a reproducible seedling system by strictly controlling the environmental factors, which we named "Ochiai method". As an example of the large-scale production of inducible plant metabolites, more than 1 g of glyceollin, an isoflavonoid-type phytoalexin, was successfully

prepared by approx. 10 kg of soybean (*Glycine max*) seedling with Koji (*Aspergillus oryzae*) inoculation. Using this stable inducible metabolite production system, molecular mechanisms of glyceollin biosynthesis were investigated by integrated transcriptomics and metabolomics approach.
Using liquid chromatography coupled with tandem mass spectrometer, more than one hundred metabolites were detected, and the metabolomic data revealed increase of amino acids and decrease of carbohydrates with the lapse of time in soybean germination. In addition, soybean phytoalexins such as glyceollin were increased by the Koji inoculation. As the genes involved in glyceollin biosynthesis were expected to be regulated synchronously at the transcriptional level, we confirmed their synchronous expression by weighted correlation network analysis (WGCNA), a clustering analysis based on Pearson correlation using transcriptome data. The result indicated that unknown transcription factors (TFs) showed significant co-expression with stress-inducible isoflavonoid biosynthetic genes. Thus, these TFs may probably function as the positive regulators in glyceollin biosynthesis.

P0778: Legumes, Soybean, Common Bean, and related
Birth and Death of a Plant Cell: Transcriptomic Analysis of the Soybean Root Hair Cells
Lise Pingault, Zhenzhen Qiao and Marc Libault, University of Oklahoma, Norman, OK
Our understanding of the genes' transcription throughout the life cycle of the plant cells is limited by the resolution of the transcriptomic resources available (plant organs or tissues *vs.* single-cell-types). Indeed, each cell-type is characterized by a specific transcriptional profile which will serve as a signature of the specific biological function of the cell-type. Selecting the soybean (*Glycine max*) root hair (RH) cells as a model to better understand the dynamic changes of gene transcription during plant cell life cycle, we performed a RNA-seq analysis on isolated RHs at 84, 96, 120, 276, 288, 312 and 336 hours after sowing (HAS) plants. We found that 84.2 % of the annotated soybean genes are expressed in at least one of the RH developmental conditions. Comparing our RH RNA-seq datasets with the transcriptomes of 7 organs [i.e., flower, leaf, nodule, apical meristem, root tip, stripped root (i.e. root devoid of RHs), and green pods], we identified 2,385 genes preferentially expressed in RHs. Among them, 95 soybean genes were repetitively characterized as preferentially expressed in RH across the entire soybean RH life cycle. The clustering of these 2,385 soybean genes according to their expression profile in RH cells revealed 915 genes preferentially expressed during the early stages of RH development (i.e. 84-HAS, 96-HAS and 120-HAS; elongation of the RH cell), whereas 1,096 genes are preferentially expressed in mature RHs (i.e. 288-HAS, 312-HAS, 336-HAS). These results support the dynamic regulation of the soybean transcriptome during the entire life cycle of the RH.

P0779: Legumes, Soybean, Common Bean, and related
Analysis of Small RNA and Gene Expression in Soybean Roots Related to *Rhg1*-Mediated SCN Resistance
Usawadee Chaiprom, University of Illinois at Urbana-Champaign, Urbana, IL, Tong Geon Lee, Gulf Coast Research and Education Center, Wimauma, FL and Matthew E. Hudson, University of Illinois, Urbana, IL
Soybean cyst nematode (SCN) is a biotrophic pathogen that invades soybean roots. The soybean yield suppression due to SCN infestation has been estimated to cost one billion US dollars per year. The majority of SCN-resistant soybean varieties are derived from a common resistance source (PI 88788) and carry the same resistance locus (*Rhg1*) which is a copy number polymorphism of 1-11 copies of a 31.2kb genomic region. Since *Rhg1*-mediated resistance is so widely used, this study aims to identify regulatory variation and differentially expressed (DE) genes in soybeans with copy number variation at the *Rhg1* locus. Analysis of sRNA and mRNA expression at the *Rhg1* locus showed higher accumulation of sRNAs and transcripts for three genes at the *Rhg1* locus in the resistant lines Peking (3 copies) and PI 88788 (9 copies) compared to their abundance in a susceptible line, W82 (1 copy). Isogenic Fayette plants were found to have no statistically significant difference in expression of genes and sRNA mapping levels at the *Rhg1* locus even though the copy number variation at the locus ranged from 9 to 11 copies. A co-expression network constructed from DE genes revealed candidate genes related to ethylene-dependent defense response and oxidation-reduction process. The Fluidigm-based qRT-PCR is used to profile temporal expression of selected candidate genes in SCN-infected roots of soybeans with copy number variation at the *Rhg1* locus.

P0780: Legumes, Soybean, Common Bean, and related
Comparative Transcriptome and Network Analyses of Photoperiodic Flowering Genes Underlying Maturity Loci (E Loci) in Soybean
Faqiang Wu[1], Minglei Wang[1], Xiaohan Kang[2], Bruce Hajek[2] and Yoshie Hanzawa[1], (1)California State University, Northridge, Northridge, CA, (2)University of Illinois at Urbana-Champaign, Urbana, IL
Soybean (*Glycine max)*, a major oil and protein source for human and livestock, flowers in response to a photoperiod change from long day to short day. Several maturity loci (*E loci*) are known to control photoperiodic flowering of soybean, but their mechanisms are largely unclear. Here we characterize soybean response to photoperiod in genome-wide gene expression patterns, with 315 pairwise comparisons of transcriptomic level using the RNA-sequencing data consisting of three different time points in a day under three different photoperiodic treatments. Six genotypes were used in this study; soybean cultivars Clark and Williams 82, and four near isogenic lines (NILs) of *E* loci (*E1*, *E2*, *E3*, and *E5*). While *E* loci exhibited effects in all conditions, we observed specific *E* loci affected gene expression width (gene numbers) and depth (gene expression levels) under specific conditions, as well as opposite directions of gene regulation (up/down), involving more than 2,000 differentially expressed genes under the control of *E* loci, including 61 flowering genes. Network modeling was performed using the 61 differentially expressed flowering genes to construct the circadian clock and flowering-time gene networks in soybean from the RNA-sequencing data under different photoperiods. Part of the regulatory network was verified by known regulatory interactions in literature or by monitoring gene expression changes in the transgenic soybeans. This study would advance our understanding of the roles of *E* loci in soybean flowering control.

P0781: Legumes, Soybean, Common Bean, and related
Genetic Dissection of Dynamic Plant Height and Number of Nodes on Main Stem in Soybean
Fangguo Chang, Nanjing agriculture university, Nanjing, China

Plant architecture in cultivated soybean has diverse types with less genetic study. Plant height (PH) and number of nodes on main stem (NN) serve as crucial traits for soybean plant-type and yield. Here, we conducted a genome-wide association study to dissect the genetic base of PH and NN at different growth stages in summer-planting soybeans. A total of 368 soybean lines were planted in two sites, and the performance of PH and NN at V5, R1, and R8 growing stages was investigated. Total 62423 single-nucleotide polymorphism markers were obtained by restriction-site associated DNA sequencing. Under the $P<1\times10^{-4}$ significant level, a total of 19 and 23 loci associated with PH and NN were identified respectively. The number of loci ranged from 2 to 5 and 5 to 6 for PH and NN at three stages respectively. Among them, two major loci on chromosome 10 and 19 overlapped with the known *E2* gene for flowering time and *Dt1* for stem growth habit. We classified the entire population into two sub-populations based on the genotypes of the highest association site of the *Dt1* and *E2* locus respectively. Additional loci associated with PH and NN were detected in comparison with the whole population. It was demonstrated there was no interaction effect between *Dt1* and *E2*. These results can help to understand the genetic architecture of PH and NN, and also provide valuable information for future ideal plant type breeding in soybean.

P0782: Legumes, Soybean, Common Bean, and related

Transcriptomics and Metabolomics of Soybean Roots Provide Insights into the GmMYB176-Mediated Regulation of Isoflavonoid Biosynthesis

Arun Kumaran Anguraj Vadivel[1,2], Justin Renaud[1], Sateesh Kagale[3] and **Sangeeta Dhaubhadel**[1,2], (1)Agriculture and Agri-Food Canada, London, ON, Canada, (2)Department of Biology, University of Western Ontario, London, ON, Canada, (3)National Research Council Canada, Saskatoon, SK, Canada

Isoflavonoids are a group of plant natural compounds synthesized almost exclusively by legumes and are abundant in soybean seeds and roots. They play important roles in plant-microbial interactions and the induction of nod gene expression in Rhizobia that form nitrogen fixing nodules on soybean roots. Isoflavonoids also contribute to the positive health effects associated with soybean consumption by humans and animals. An R1MYB transcription factor GmMYB176 regulates isoflavonoid biosynthesis by activating *chalcone synthase* (C*HS*)*8* gene expression in soybean. Using a combination of transcriptomic and metabolomic analysis of GmMYB176 RNAi silenced (GmMYB176-Si) and GmMYB176 overexpressed (GmMYB176-OE) soybean hairy roots, we identified a total of 75 differentially expressed genes (DEGs) and 995 differentially produced metabolites in GmMYB176-Si hairy roots, and 7099 DEGs and 149 differentially produced metabolites in GmMYB176-OE hairy roots. By a targeted approach, 11 isoflavonoid biosynthetic genes and 6 metabolites were identified as differentially regulated in GmMYB176-OE and GmMYB176-Si soybean hairy roots. Taken together, our results demonstrate the complexity of isoflavonoid biosynthesis in soybean roots, and suggest that a coordinated expression of pathway genes, substrate flux and product threshold level contribute to the dynamic of the pathway regulation.

P0783: Legumes, Soybean, Common Bean, and related

A Zinc Finger Transcription Factor GmZF-1, Regulates Isoflavone Biosynthesis in Soybean

Junming Sun[1], Bin Li[1] and Shengrui Zhang[2], (1)Institute of Crop Sciences,CAAS, Beijing, China, (2)Institute of Crop Sciences,CAAS, China

Soybean isoflavone belongs to a group of secondary metabolites derived from the phenylpropanoid pathway and are mainly produced in legumes. Due to their important roles for plants and human health, studies on accumulation of isoflavone have been performed worldwide. The ultimate goal of the studies is to illustrate the genetic basis of the isoflavone accumulation and to develop a series of cultivars with varying isoflavone content.

In this study, we constructed a high-density soybean genetic map using RILs by specific length amplified fragment sequencing (SLAF-seq) method. The map consisted of 5,785 SLAFs on 20 linkage groups (LGs) and spanned 2,255.18 cM in genome size with an average distance of 0.43 cM between adjacent markers. Based on this map, we identified 41 QTL that contributed to the isoflavone content. Moreover, 11 of these 41 QTL (including six novel loci) were associated with isoflavone content across multiple environments. One of them, *q*IF20-2, contributed to a majority of isoflavone components across various environments and explained a high amount of phenotypic variance (8.7% - 35.3%).

The fine mapping and candidate gene screening of major QTL(*q*IF20-2) were carried out by the bulk segregant analysis sequence (BSA-seq) and the secondary segregation population encryption. The results indicated that a key gene (*GmZF-1*) controlling soybean isoflavone accumulation was found. The functional analysis showed that the zinc finger transcription factor encoded by *GmZF-1* was a positive regulator in the soybean isoflavone accumulation.

P0784: Legumes, Soybean, Common Bean, and related

Transcriptomic Analysis of Fine Mapped Soybean Aphid (*Aphis glycines*) Resistance Gene, *rag1c* in Soybean (*Glycine max* L.)

Jiazheng (John) Yuan[1], Dechun Wang[2], Yong-Qiang (Charles) An[3], Zixiang Wen[2] and Umesh Rasyara[4], (1)Fayetteville State University, Fayetteville, NC, (2)Michigan State University, East Lansing, MI, (3)USDA/ARS PGRU, Saint Louis, MO, (4)MSU

The soybean aphid (Aphis glycines Matsumura) is one of the major pests of soybean [Glycine max (L.) Merr.] in the United States and Eastern Canada. More than four alleles/genes, Rag1, Rag, rag1c, and rag1b against soybean aphids have been identified in soybean germplasms Dowling, Jackson, PI567541B, and PI567598B, respectively and mapped to the same locus on chromosome 7. These resistance alleles/genes from soybean plants effectively provide resistance against different soybean aphid biotypes. The transcriptomic variation among candidate genes in rag1c interval was distinguishable between soybean aphid resistant lines and susceptible lines after soybean aphid infestation in a temporal investigation. Our evidences of the recombination breakpoints that caused the phenotypic variation of aphid resistance in the fine mapping populations and transcriptomic gene expression analysis suggest that the rag1c gene of soybean aphid resistance gene was defines into a 96 kbp interval on soybean genome flanked by KASP® SNP marker MSUSNP7-32 and MSUSNP7-20 on soybean genome (Williams 82 a2.v1).

P0785: Legumes, Soybean, Common Bean, and related

Optimization of *Agrobacterium*-Mediated Transformation in Soybean

Shuxuan Li[1], Yahui Cong[1], Yaping Liu[1], Tingting Wang[1], Qin Shuai[1], Nana Chen[1], Junyi Gai[2] and Yan Li[3], (1)Nanjing Agricultural University, Nanjing, China, (2)Soybean Research Institute, National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean (Ministry of Agriculture), State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, (3)National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China

High transformation efficiency is a prerequisite for study of gene function and molecular breeding. Agrobacterium tumefaciens-mediated transformation is a preferred method in many plants. However, the transformation efficiency in soybean is still low. The objective of this study is to optimize Agrobacterium-mediated transformation in soybean by improving the infection efficiency of Agrobacterium and regeneration efficiency of explants. The results showed that an infection efficiency of over 96% was achieved by collecting the Agrobacterium at a concentration of OD650 = 0.6, then using an Agrobacterium suspension medium containing 154.2 mg/L dithiothreitol to infect the half-seed cotyledonary explants (from mature seeds imbibed for 1 day), and co-cultured them for 5 days. On the other hand, the rates of shoot elongation were compared among six different concentration combinations of gibberellic acid (GA3) and indole-3-acetic acid(IAA). The shoot elongation rate of 34 and 26% was achieved when using the combination of 1.0 mg/L GA3 and 0.1 mg/L IAA for Jack Purple and Tianlong 1, respectively. This rate was higher than the other five concentration combinations of GA3 and IAA, with an 18 and 11% increase over the original laboratory protocol (a combination of 0.5 mg/L GA3 and 0.1 mg/L IAA), respectively. The transformation efficiency was 7 and 10% for Jack Purple and Tianlong 1 at this optimized hormone concentration combination, respectively, which was 2 and 6% higher than the original protocol, respectively.

P0786: Legumes, Soybean, Common Bean, and related

Development of SSR Markers for Genetic Diversity and Phylogenetic Studies of *Phomopsis longicolla* causing Phomopsis Seed Decay in Soybean

Shuxian Li, USDA-ARS, Crop Genetics Research Unit, Stoneville, MS and Qijian Song, USDA-ARS, Soybean Genomics and Improvement Lab, Beltsville, MD

Phomopsis longicolla T. W. Hobbs (syn. *Diaporthe longicolla*) is the primary cause of Phomopsis seed decay (PSD) in soybean, *Glycine max* (L.) Merrill. The genome of *P. longicolla* type strain TWH P74 represents one of the important fungal pathogens in the *Diaporthe-Phomopsis* complex. In this study, the draft genome sequence of *P. longicolla* type strain TWH P74 was *de novo* assembled. The *P. longicolla* genome sequence provides molecular resources for developing genetic markers for *P. longicolla*. The draft genome size was approximately 64 Mb. We examined the simple sequence repeat (SSR) in the genome, and identified 12,624 SSRs with di-, tri-, and tetranucleotide repeats of five or more in the TWH P74 whole genome sequence, which included 1,972 SSRs consisting of repeat units of di- (≥9) (919), tri- (≥8) (369), and tetranucleotide (≥7) (684). Among the 1,972 SSRs, (AT)n, (ATT)n and (AAAT)n were the most abundant motifs among di-, tri-, and tetranucleotide SSRs, respectively. The SSR markers developed from whole genome sequence will be used to analyze the genetic diversity among *P. longicolla* isolates collected from different geographical origins, as well as their phylogenetic relationships.

P0787: Legumes, Soybean, Common Bean, and related

Using High Throughput Molecular Data Combined with Phenotypic Evaluation to Identify Genetic Basis for Partial Resistance to White Mold in Soybean

André Luiz de Lima Passianotto[1], Francois Belzile[2] and Istvan Rajcan[1], (1)University of Guelph, Guelph, ON, Canada, (2)Laval University, Quebec, QC, Canada

Soybean [Glycine max (L.) Merr.] crop is frequently exposed to field pathogens including *Sclerotinia sclerotiorum* the causal agent of white mold (WM). Soybean growing regions of northern United States and Canada being cool and moist have favorable environmental conditions for the development of WM, which can cause significant damage to yield and grain quality in some years. In soybean, there is no evidence of complete resistance against this pathogen, however, partial resistance has been reported in a number of independent studies, which have rarely been validated. Genomics tools such as QTL analysis and RNA-Seq offer an opportunity to dissect the complex nature of underlying genetics for partial resistance to WM in soybean across different plant populations. The objective of this study was to use QTL mapping, RNA–*Seq* combined with GBS and phenotyping to elucidate genomic regions controlling partial resistance to WM in different genotypes. Fourteen plant introductions (PI) and 20 Canadian cultivars with known partial resistance and susceptible phenotypes were genotyped to obtain haplotypes and phenotyped using a reproducible inoculation method to determine their reaction to the pathogen. Resistance was assessed by measuring the length of the lesion on the main stem after seven days post-inoculation, which ranged between 0.9 and 19.1 cm across the genotypes. Haplotypes will be compared against the phenotype for reported QTL alleles and candidate genes variation as determined by RNA-Seq.

P0788: Legumes, Soybean, Common Bean, and related

Applying Genomics to Plant Breeding: Leveraging RNA-Seq to Characterize the *Resistance to Brown Stem Rot 3* (*Rbs3*) Locus and Downstream Signaling Networks in Soybean

Chantal McCabe, United States Department of Agriculture, Ames, IA and Michelle A. Graham, USDA-ARS-MWA-CICGRU, Ames, IA

Crop damage from pests and pathogens results in a loss of over $60 billion in the United States each year. Brown stem rot (BSR), caused by the fungus *Phialophora gregata*, reduces soybean yield by up to 38%. Three dominant BSR resistance loci have been identified: *Rbs1*-*Rbs3*, however the genes and gene networks regulating resistance remain unknown. Further, identifying resistant germplasm by genotyping or phenotyping remains difficult due to complexities of the soybean/*P. gregata* interactions. Therefore, we conducted RNA-Seq of *P. gregata* infected and mock-infected leaf, stem, and root tissues of an *Rbs3* resistant soybean genotype (PI 437970). Tissue samples were collected 12, 24, and 36 hours after treatment. Our analyses identified genes differentially expressed in response to infection at each time point. Little overlap was identified in differential gene expression between tissues, suggesting each tissue has a distinct response to *P. gregata* infection. In leaves, the strongest gene expression occurred 12 and 24 hours after infection. These differentially expressed genes primarily functioned in cell

proliferation and organ development. In stems and roots, the strongest gene expression occurred 36 hours after infection and had functions associated with defense response. These results will provide additional information about mechanisms of BSR resistance. With this knowledge we can develop markers to screen lines for resistance earlier than phenotyping allows. Additionally, VIGS will be used to characterize candidate resistance genes and downstream defense responses.

P0789: Legumes, Soybean, Common Bean, and related

Identification of a QTL for Resistance to *Pythium ultimum* var. *ultimum* in a Soybean RIL Family

Kelsey Scott[1], Christine Balk[1], Jonell Winger[1], Leah McHale[2] and Anne E. Dorrance[1], (1)The Ohio State University/OARDC, Wooster, OH, (2)The Ohio State University, Columbus, OH

Pythium seed and root rot of soybean can cause severe stand losses, particularly under saturated growing conditions. The soilborne oomycete *Pythium ultimum* var. *ultimum* is one of the most prevalent and pathogenic *Pythium* species that contributes to disease development in Ohio. Several soybean Nested Association Mapping (NAM) families have been identified as segregating for resistance to key Ohio oomycete species, including *Pythium ultimum* var. *ultimum*. Our objective was to identify the quantitative trait loci (QTL) that confer resistance to *P. ultimum* var. *ultimum* in a F_5 RIL family derived from the cross HS6-3976 x IA3023, which is part of the larger NAM population. The NAM family was evaluated for resistance to *P. ultimum* var. *ultimum* isolates using a greenhouse cup assay. The disease response of the NAM family was indicative of a quantitative trait; the mean root rot score ranged from 1 to 5 with a mean score of 3. Composite interval mapping was performed using genome-wide logarithm of odds thresholds of 3.2 (type I error rate of $\alpha = 0.5$). Preliminary screening identified one QTL on chromosome 2 associated with root weight. This QTL is a potentially useful source of quantitative resistance to *P. ultimum* var. *ultimum*.

P0790: Legumes, Soybean, Common Bean, and related

Identification of Molecular Mechanisms Underpinning Partial Resistance to *Phytophthora sojae i*n *Glycine max* using a Systems Genetics Approach

Cassidy Gedling[1], Saranga Wijeratne[2], Stephanie Karhoff[3], Anne E. Dorrance[1] and Leah K. McHale[3], (1)The Ohio State University/OARDC, Wooster, OH, (2)The Ohio State University, Wooster, OH, (3)The Ohio State University, Columbus, OH

Phytophthora sojae is a leading pathogen of soybean, causing root and stem rot (PRR) across the Midwestern U.S. Race-specific R-genes are commonly deployed to manage *P. sojae*. However, due to a large number of pathotypes, partial resistance is important in several regions and thought to put less selection pressure on *P. sojae* populations. Twenty phenotypic quantitative trait loci (pQTL) for partial resistance against *P. sojae* were previously mapped in a F9:11 Conrad x Sloan recombinant inbred line (RIL) population on chromosomes 1, 4, 9, 15, 16, 18, and 19. However, these regions encompass large portions of the genome. Thus, the overall goal of this study is to identify molecular mechanisms contributing towards partial resistance, with the specific objective of reducing the list of candidate genes potentially underpinning pQTL. A systems genetics approach that incorporates expression QTL (eQTL) mapping, functional genomics, and gene co-expression analysis was taken to achieve this goal. A subset of 93 RILs from the Conrad x Sloan population were inoculated with isolate 1.S.1.1 using the tray-test method, tissue at the inoculation site was collected 24 hours after inoculation from mock and inoculated samples, RNA was extracted, and sequenced using Illumina Hi-seq. A greater number of eQTL were mapped in inoculated samples relative to mock, indicating transcriptional reprogramming due to *P. sojae* infection. Of the 25 co-expression modules identified, one was positively correlated with disease and one negatively. Meta-analysis of eQTL/pQTL are in progress. Putative genes regulating co-expression networks will be validated in our functional gene analysis pipeline.

P0791: Legumes, Soybean, Common Bean, and related

Molecular Identification of *Rps11*, a Gene Resistant to Soybean *Phytophthora sojae*

Weidong Wang, Liyang Chen, Jieqing Ping, Josh Fitzgerald and Jianxin Ma, Purdue University, West Lafayette, IN

Phytophthora root and stem rot (PRSR) is a major disease in soybean (*Glycine max* (L.) Merr.), and causes massive worldwide yield loss every year. The most effective and environmentally friendly management strategy is deployment of race-specific resistant soybean cultivars. The objective of this study is to identify and clone novel Rps (Resistance to P. sojae) gene conferred in soybean landrace PI94527. We demonstrated that a single locus, denoted as *Rps11*, is responsible to the resistance to all the isolates we used in this study, indicating a broad resistance spectrum. In addition, *Rps11* was mapped to 61 kb genomic region defined by simple sequence repeat (SSR) marker BARCSOYSSR_07_0295 and insertion/deletion marker InDel_2, according to the soybean reference genome, using a large mapping population including 2640 F3 individuals. Five genes were predicted in this interval based on the reference genome, including a gene coding a NB-ARC domain containing disease resistance protein, which is considered to be the best candidate gene so far. Future work including finer mapping from a larger population comprising 7680 F4 individuals, validating the function of the candidate gene using technologies such as qPCR, QTL_seq as well as CRISPR/CAS 9.

P0792: Legumes, Soybean, Common Bean, and related

Genomic Regions Responsible for Soybean Cyst Nematode Resistance Revealed in a Genome-Wide Association Study

Dung Thuy Tran[1], James Noe[2], Prakash Arelli[3] and Zenglu Li[1], (1)Institute of Plant Breeding, Genetics & Genomics and Department of Crop and Soil Sciences, The University of Georgia, Athens, GA, (2)Department of Plant Pathology, The University of Georgia, Athens, GA, (3)USDA-ARS-SEA, Jackson, TN

Soybean cyst nematode (SCN) is the most destructive pest affecting soybeans in the USA. Two major resistant loci, *Rhg1* and *Rhg4* on chromosome (Chrs)18 and 8 have been reported in many SCN resistant sources.Of those, PI88788(*rhg1*) and 'Peking'(*rhg1*/*Rhg4*) have been widely used to develop resistant cultivars in the USA.However, it is essential to identify new sources of resistance before SCN overcomes these two resistant sources.We screened 462 soybean accessions using greenhouse screenings and genotyped them with three functional SNPs developed at *Rhg1* and *Rhg4*.Of 462 accessions, 50 were classified as the 'Peking'type resistance through *Rhg1* and *Rhg4* loci, while 30 classified as the PI88788 type resistance with *Rhg1*. Fifty- eight accessions that were rated as SCN resistance in greenhouse phenotyping do not carry either 'Peking'or PI88788 resistant alleles at *Rhg1* and *Rhg4* loci. Based on haplotype analysis at these loci assembled with the

Soy50KSNP Infinium Chip, these lines were grouped separately from 'Peking' and PI88788. The result indicated these 58 accessions might possess novel SCN resistant genes or alleles. The genome-wide association study was performed on this panel of 462 accessions using >35,000 SNPs from Soy50KSNP Infinium Chip.Thirteen SNPs on Chrs.2, 7, 8, 10 and 18 were significantly associated with SCN resistance. Three of 13 SNPs were located at two known major QTLs: *Rhg1* and *Rhg4* on Chrs.18 and 8.The identified SNPs and candidate genes from this study might be beneficial for developing markers to be used for marker-assisted breeding and developing soybean cultivars with novel sources of SCN resistance.

P0793: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Study (GWAS) of Soybean Cyst Nematode Resistance in Soybean

Jun Qin[1], Ainong Shi[1] and Senyu Chen[2], (1)University of Arkansas, Fayetteville, AR, (2)Department of Plant Pathology, University of Minnesota, St. Paul, MN

Soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) is the most serious yield-limiting pathogen on soybean (*Glycine max* (L.) Merr.) and use of the host genetic resistance is the most effective and environmental friendly method to control it. The objective of this study was to identify quantitative trail locus (QTL) conditioning SCN resistance and to develop single nucleotide polymorphisms (SNP) markers tagging the resistance in soybean. A total of 274 accessions with high levels of resistance and susceptibility to SCN from the USDA Soybean Germplasm Collection were selected and phenotyped with SCN HG Type 0 (race 3). A genome-wide association study (GWAS) was conducted to identify QTL controlling SCN resistance based on 29,383 SNPs from the SoySNP50K iSelect SNP beadchip. We identified three major QTLs, i.e. SCN-qtl7, SCN-qtl11, and SCN-qtl18, which were on chromosome (chr.) 7, chr. 11, and chr. 18, respectively. Seven SNPs in SCN-qtl7, five SNPs in SCN-qtl11, and six SNPs in SCN-qtl18 regions were strongly associated with the resistance to race 3. The highest LOD value of the SNPs at the SCN-qtl7 QTL was 50.8, 33.3, and 14.8 based on the analysis methods of single marker regression (SMR), general linear model (GLM), and mixed linear model (MLM), respectively. The R-squared value of this SNPs was 57.2%, 30.3%, and 26.9%, respectively, based on the three methods. The highest LOD value of the SNPs at SCN-qtl11 was 36.2, 20.3, and 8.4, and their R-squared value was 45.6%, 20.6%, and 13.6% based on SMR, GLM, and MLM, respectively. The highest LOD value of the SNPs at SCN-qtl18 was 31.2, 16.6, and 7.4 with R-squared value of 40.0%, 16.8%, and 11.9% based on SMR, GLM, and MLM, respectively. The identified QTLs and SNP markers will assist breeders to improve SCN resistance in soybean.

P0794: Legumes, Soybean, Common Bean, and related

Identification and Evaluation of Quantitative Trait Loci Associated with Soybean Cyst Nematode Resistance in PI437654-Derived Populations

Xin Lu, University of Guelph, Guelph, ON, Canada

Soybean Cyst nematode (SCN) is the most damaging pest of soybean worldwide. Damage by SCN costs North American soybean growers over $1 billion each year. Growing resistant cultivars together with crop rotation is the most effective way to control SCN; however, using diverse sources of resistance is also recommended to prevent resistance breakdown. To date, about 98% of SCN-resistant cultivars in North America are derived from the soybean accession PI 88788 and breakdown of 88788-derived resistance is ascertained in some areas. Developing new SCN-resistant cultivars with different resistant genes may reduce the risk of genetic shift in SCN populations. The main objective of this study was to detect quantitative trait loci (QTL) associated SCN genes in PI 437654-derived soybeans. Two F_{4-5} recombinant inbred line (RIL) populations derived from crosses between two high yielding Ontario-adapted SCN-susceptible soybeans, OAC Brooke and OAC Calypso, and LD07-3419, which carries its SCN genes from PI 437654, were used for this study. In total, 347 RILs were evaluated for resistance to SCN (*Heterodera glycines*) HG type 2.5.7 in a greenhouse and genotyped using genotyping-by-sequencing method. Three QTL on Chr.9, Chr.12 and Chr.18 were identified associated with the resistance. . These SCN-related markers/QTL will facilitate the development of SCN-resistant soybeans through marker-assisted-selection.

P0795: Legumes, Soybean, Common Bean, and related

A Trio of Evolutionary Mechanisms Give Rise to a Diversity of Parasitic Effector Genes in the Soybean Cyst Nematode

Rick Masonbrink[1], Thomas Maier[1], Arun S. Seetharam[1], Etienne Lord[2], Parijat S. Juvale[1], Benjamin Mimee[2], Sebastian Eves-van den Akker[3], Andrew J. Severin[1] and Thomas J. Baum[1], (1)Iowa State University, Ames, IA, (2)Agriculture and Agri-Food Canada, Gouin, QC, Canada, (3)University of Dundee, Invergowrie, United Kingdom

Heterodera glycines or the soybean cyst nematode (SCN), an obligatory and sedentary endoparasite causes over a billion dollar yield loss every year in USA alone. The costs and environmental hazards of current nematicides, and the emergence and spread of SCN populations that overcome the limited available sources of natural resistance are exacerbating the problem. Since effectors, the proteins secreted by SCN into host root tissues are thought to play crucial roles in overcoming host defenses understanding the mechanisms of effector acquisition, diversification and selection is critical in improving SCN management practices. Toward that end, we present a draft genome assembly of the SCN. Using PacBio long read technology, we assembled and annotated 738 contigs into 123Mb with 28,273 genes. The genome contains significant numbers of repeats (34%), tandem duplicates (10.5%) and horizontal gene transfer events (156 genes). Using this genome we explored potential mechanisms for how effectors originate, duplicate, and diversify. Specifically, we found that horizontal gene transfer events are frequently associated with known effectors, tandem duplications are abundant and have greater proportions of effectors than non-duplicated regions, and a large proportion of genes in the genome overlap with repetitive elements. We hypothesize that horizontal gene transfer events, tandem duplications and transposable elements are involved in acquisition and diversification of effector genes. In addition, we hypothesize that the genetic diversity of SCN populations contributes to the rapid adaptability of SCN to resistant soybean lines through natural selection.

P0796: Legumes, Soybean, Common Bean, and related

Positional Gene Cloning and Sequencing of Two Soybean Cyst Nematode Resistance Loci

Christina Fliege, University of Illinois, Urbana, IL

Glycine soja, the wild annual ancestor of soybean, can be used as a source of genetic diversity for the crop. PI 468916, a Glycine soja line, contains two resistance QTL (cq-SCN006 and cq-SCN007) to soybean cyst nematode (SCN) Heterodern glycines, a major pest of soybean. These loci have been fine mapped to intervals of 300 kilobase pairs and 147 kilobase pairs, respectively. Through screening a fosmid library and de novo Illumina sequencing, the sequences of the two intervals are being elucidated in detail. The alignment of the sequence of the DNA from the resistance gene interval in PI 468916 to Williams 82 allows for the discovery of differences between the two sequences, of which there are many. Changes in the sequences include single nucleotide variants as well as insertion or deletions. In some areas of these loci we have observed a very high frequency of variants (both SNVs and indels) which are similar to a kataegis or mutational storm. These regions of localized hypermutation have been observed in some cancer genomes. We analyzed these differences in the sequences of PI 468916 and Williams 82 for their potential to alter encoded proteins, and affect splice sites or promoters. Very many of these differences could be the source of the SCN resistance of these loci. Precise identification of these novel resistance loci could be helpful in precision breeding approaches or using transgenic methods to introduce new resistance into soybean.

P0797: Legumes, Soybean, Common Bean, and related

Molecular Characterization of a Candidate Rsv3 Gene from a Soybean Genotype that Confers Strain-Specific Resistance to Soybean Mosaic Virus

Phu-Tri Tran, Seoul National University, Seoul, South Korea, Kristin Widyasari, Seoul National University, SEOUL, Korea, Republic of (South), Jang-Kyun Seo, Seoul National University, Pyeongchang, Korea, Republic of (South) and **Kook-Hyung Kim**, Seoul National University, Seoul, Korea, Republic of (South)

Soybean mosaic virus (SMV), a member of the genus *Potyvirus*, significantly reduces soybean production worldwide. *Rsv3,* which confers strain-specific resistance to SMV, was previously mapped between the markers A519F/R and M3Satt in chromosome 14 of the soybean [*Glycine max* (L.) Merr.] genotype L29. Analysis of the soybean genome database revealed that five different NBS-LRR sequences exist between the flanking markers. Among these candidate *Rsv3* genes, the full-length cDNA of the *Glyma.14g204700* was successfully cloned from L29. Over-expression of *Glyma.14g204700* in leaves inoculated with SMV inhibited viral infection in a soybean genotype lacking *Rsv3*. In addition, the transient silencing of the candidate gene caused a high accumulation of an avirulent strain in L29 carrying *Rsv3*. Our results therefore provide additional line of evidence to support that *Glyma.14g204700* is likely *Rsv3* gene that confers strain-specific resistance to SMV.

P0798: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Study Identifies *Bc-U* and *Bc-1^2* candidate Genes Conferring Resistance to Bean Common Mosaic Virus

Alvaro Soler[1], John Hart[2], Timothy Porch[2], Phillip McClean[3] and Phillip Miklas[4], (1)USDA-ARS, PROSSER, WA, (2)USDA-ARS Tropical Agriculture Research Station, Mayaguez, PR, (3)Dept of Plant Sciences, North Dakota State University, Fargo, ND, (4)USDA-ARS, Prosser, WA

Bean common mosaic virus (BCMV) and *Bean common mosaic necrosis virus* (BCMNV), members of the genus Potyvirus, cause a globally prevalent and devastating seed-borne disease of common bean (*Phaseolus vulgaris* L.). Host resistance conferred by combinations of the dominant *I* gene and one or more of six recessive alleles *bc-u*, *bc-1*, *bc-1^2*, *bc-2*, *bc-2^2* and *bc-3* interacts with strains across eight pathogroups (PG). Markers for *I*, *bc-1^2*, and *bc-3* genes were developed, and candidate genes discovered for *I* and *bc-3*. Our objective was to conduct fine mapping toward candidate gene analysis of the *bc-1^2* gene. Two association mapping populations, the Durango Diversity Panel (DDP) and Snap Bean Association Panel (SnAP) consisting of 182 and 376 accessions, respectively, were screened in the greenhouse for separate reactions to NL-3 (PG 6) and NL-8 (PG 3) strains of BCMNV. Whole genome sequencing and genotype-by-sequencing were used to identify SNPs in the DDP (687,782 SNPs) and SnAP (23,304 SNPs), respectively. GWAS revealed two peaks on chromosomes Pv03 and Pv05, corresponding with *bc-1^2* and putative *bc-u* genes, respectively. Missense variants were detected in candidate genes sequenced by Sanger sequencing. A set of SNP markers were developed and screened by Tm-shift genotyping in this study, providing an important tool for marker-assisted selection programs aimed at improvement of common bean cultivars.

P0799: Legumes, Soybean, Common Bean, and related

Using Sparse Yield Trial Data for GWAS: Insights from the Cooperative Dry Bean Nursery

Alice H. MacQueen, University of Texas at Austin, Austin, TX, Jeffrey W. White, USDA-ARS, Maricopa, AZ, Phillip McClean, Dept of Plant Sciences, North Dakota State University, Fargo, ND and Tom Juenger, The University of Texas at Austin, Integrative Biology, Austin, TX

The Cooperative Dry Bean Nursery (CDBN) is an ongoing 60+ year collaboration across the United States and Canada. CDBN collaborators have collected phenotypic data for important agronomic traits for over 500 common bean varieties grown in 84 locations. When combined with genomic data, this major phenotyping effort offers unparalleled opportunities to determine how genetic factors affect both phenotypic variation and genotype by environment interactions (GxE). However, the incomplete nature of this dataset – less than 2% of all possible year by location by variety combinations are present– poses a challenge for traditional genome-wide association mapping (GWAS) analyses.

In collaboration with current common bean sequencing efforts, we sequenced 314 varieties from the CDBN and established a GWAS panel. We describe three methods to reduce the dimensionality of this dataset to allow GWAS. For these methods, we determined the average phenotypic response and the variation in the phenotypic response for four traits. First, we considered phenotypes as factors of one of seven geographic clusters and of common bean variety. Second, we use Finlay-Wilkinson joint regression analysis to regress variety phenotypes against the average phenotypes within the seven geographic regions. Third, we fit environmental gradient models using weather data for each location and year as predictors for phenotypes within the seven geographic regions. Our results suggest that there is substantial GxE in the CDBN, as most genomic regions with significant associations only influenced a subset of geographic regions. We discuss genomic regions with pleiotropy, or effects on multiple phenotypes.

P0800: Legumes, Soybean, Common Bean, and related

Bean Adapt: The Genomics of Adaptation during Crop Expansion of Common Bean

Roberto Papa, Università Politecnica delle Marche, Ancona, Italy, Alisdair R. Fernie, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany, Paul Gepts, University of California, Davis, CA, Andreas Graner, Leibniz Institute of Plant Genetica and Crop Plant Reserch, Gatersleben, Germany and Scott A. Jackson, University of Georgia, Athens, GA

Domesticated species spread outside their native range and faced new environmental challenges. This evolutionary scenario makes crops excellent models to deeply understand adaptation to new ecological conditions. The BEAN_ADAPT project, funded through the 2ndERA-CAPS call, ERA-NET for Coordinating Action in Plant Sciences, aims to dissect out the genetic basis and phenotypic consequences of the adaptation to new environments of the common bean (*Phaseolus vulgaris* L.), through the study of the introduction, from the centres of domestication in the Americas, and the expansion through Europe, as a recent and historically well-defined event of rapid adaptation.

We re-sequenced 220 American and European common bean accessions at an average 8X coverage per accessions. We identified ~1.5 million of SNPs with less than 5% of missing data and analysed them to characterize genome-wide variation. Moreover, this sample represents a subsample of a nested collection of 500 common bean accessions which was phenotyped both in controlled conditions and in multi-field trials at two different latitudes (Northern Germany and Southern Italy).

Here we present the results of genomic data based on population genomics approaches and of integration of genomic, transcriptomic, metabolomic and phenotypic data to identify genes and/or genomic regions associated to important traits related to adaptation of *P. vulgaris* to the European agroecosystems.

P0801: Legumes, Soybean, Common Bean, and related

Inferring the Gene Regulatory Network of Soybean and Common Bean *Nodule Inception 1* (*NIN1*) Genes.

Prince Zogli[1], Lise Pingault[1], Zhenzhen Qiao[1], Peter Klein[2] and Marc Libault[1], (1)University of Oklahoma, Norman, OK, (2)Kansas State university, Manhattan, KS

Medicago truncatula Nodule Inception 1 (*MtNIN1*) encodes a transcription factor playing a major role during the nodulation process, a mutualistic symbiotic interaction between plant roots and soil bacteria from the gender *Rhizobia*. To date, similarly to *MtNIN1*, the *Casuarina glauca* and *Lotus japonicus NIN1* genes have also been characterized as major regulators of the nodulation process. However, the exact function of NIN1-like proteins in *Glycine max* (soybean) and *Phaseolus vulgaris* (common bean) is unknown.

In this study, we first used bioinformatics tools to identify 4 and 2 *NIN1*-like genes in soybean and common bean, respectively. Among them, 2 soybean and one common bean genes belong to the same clade than *MtNIN1*. RNA-Seq analyses of the expression level of these soybean and common bean *NIN1s* revealed their high transcriptional activity in *Rhizobia*-inoculated root hairs and nodules. This result was further confirmed upon expression in plant roots of transcriptional fusions between the promoter sequences of the soybean and common bean *NIN1s* and the *GUS* reporter gene. The similarity of the transcriptional pattern of the legume *NIN1s* suggest the conservation of their function between legume species including, potentially, the gene networks under their control. To confirm this hypothesis, chromatin immunoprecipitation followed by sequencing (ChIP-Seq) and protein pull-down assays are currently underway to infer the Medicago, soybean and common bean NIN1s regulatory networks as well as protein partners. Taken together, our preliminary data provide valuable information for the co-evolution of *NIN1* gene family between legumes.

P0802: Legumes, Soybean, Common Bean, and related

Identification of Loci Associated with Rapid Early-Season Growth in Common Bean.

Travis Allen Parker, UC Davis Department Of Plant Sciences, Davis, CA, Antonia Palkovic, Department of Plant Science, Davis, CA and Paul Gepts, University of California, Davis, CA

Weeds are a major burden in most crop ecosystems. Crop plants require vigorous early season growth rate to compete effectively against weeds. Improving this trait is a major objective for plant breeders. Unfortunately, the genetic basis of early season vigor has historically been poorly understood, leaving breeders without the tools needed for efficient marker-assisted selection (MAS). Using the common bean Middle American Diversity Panel (MDP) and a high-throughput UAV-based data acquisition and analysis pipeline, we identified loci on four chromosomes that are significantly correlated with high early-season canopy growth rate. Ongoing projects based on 2017 data include analysis of growth rate in a recombinant inbred population derived from a cross between extremely vigorous and extremely non-vigorous parents, as well as correlations between growth rate and indices based on additional spectral data, such as canopy temperature and NDVI. If confirmed, the loci identified in these studies could be used in MAS to improve weed competitiveness of common bean.

P0803: Legumes, Soybean, Common Bean, and related

Meta-QTL Analysis in Common Bean to Uncover the Genetic Architecture of Iron and Zinc Concentration in Seed.

Paulo Izquierdo, Michigan State University, East Lansing, MI, Carolina Astudillo, Crop Science Divistion, West Sacramento, CA, Asif M. Iqbal Qureshi, Department of Plant, Soil & Microbial Sciences, Michigan State University, East Lansing, 48823, EAST LANSING, MI, Matthew W. Blair, Department of Agricultural and Environmental Sciences, Tennessee State University, Nashville, TN, Bodo Raatz, International Center for Tropical Agriculture (CIAT), Cali, Colombia and Karen Cichy, USDA-ARS & Department of Plant Soil & Microbial Sciences, MSU, East Lansing, MI

In the last years have been an increasing interest to produce food with higher nutritional quality around the world, especially in countries with nutritional deficiencies. Common bean (Phaseolus vulgaris L.) is the most important legume for human consumption worldwide and due to its high concentration of iron and zinc is use in biofortification programs around the world. Several studies have been developed in common bean to detect genomic regions associated with iron and zinc concentration in seed, but unfortunately by the complexity of these traits, has not been possible design markers to marker assisted selection. The goal of this study was to use all the QTLs for iron and zinc concentration published up to date to conduct a meta-analysis of these traits in common bean. In total, we integrated seven QTL studies in Andean and Middle American intra and inter genepool populations to identify the regions in the genome that control iron and zinc accumulation in seeds. In total, 4 Meta-QTL specific to Fe and 3 Meta QTL specific to Zn were identified. Additionally, eight Meta QTL that co-localized for Fe and Zn

concentration were identified across 8 chromosomes. In the 15 Meta-QTLs we identified 15 candidate genes that belong to seven gene families that have been related with transport of iron and zinc in plants.

P0804: Legumes, Soybean, Common Bean, and related

Transcriptomic Analysis of Common Bean Grown Under Salinity Stress

Mariama T Brown[1], Adrianne Brown[2], Grant Blank[1], Dennis McIntosh[1], Mayavan Subramani[1], Rita Hayford[3], Mollee C Dworkin[1], Antonette Todd[1] and Venu (Kal) Kalavacharla[1], (1)Delaware State University, Dover, DE, (2)Tuskegee University, Tuskegee, AL, (3)University of Delaware, Newark, DE

Common bean (*Phaseolus vulgaris* L.*)* is the most important consumable food legume globally and is a major source of protein in the diets of many people. Numerous farmers depend upon the cultivation of this crop as a source of household income. However, production of common bean is affected by environmental stressors including unfavourable soil conditions such as salinity that creates a critical threat to agriculture. The result of surplus minerals such as Na^+ and/or Cl^- on plant growth is known as salt stress, which limits water and nutrient availability needed for plant development. Common bean yields are reduced up to 20% at 10mM of NaCl. Plants adapt to and survive stress factors by activating stress responses through genetic pathways that cause physiological changes. Common bean has the ability of expressing or suppressing stress responsive genes to help cope under adverse stressors. In this study, RT-qPCR and RNA-seq will be used to identify differential expression of some salt genes (LEA18, HSP90 and ERD1). This will be done by spatial (leaf, root, stem, flower) and temporal (10 days and 5 weeks salt stressed) analysis in the Meso-American common bean cultivar "Sierra". The plants were salt stressed under hydroponic conditions of 0 (control), 50 and 150 mM NaCl. The data generated in this study will elicit possible candidate genes responsible for salt stress and will serve as a genetic resource for development of salt tolerance in bean. This will also ultimately contribute useful tools for plant breeders towards genetic improvement of crops.

P0805: Legumes, Soybean, Common Bean, and related

Recognizing the Effects of Salinity on Root Rot Pathogens of Common Bean (*Phaseolus vulgaris* L.) and Exploring Differences in Host Gene Expression

Dylan C Lynch[1], Rubella S Goswami[2], Joey Cainong[1] and Venu (Kal) Kalavacharla[1], (1)Delaware State University, Dover, DE, (2)USDA-National Institute of Food and Agriculture, Washington, DC

Common bean (*Phaseolus vulgaris* L.) is produced worldwide as a vital food crop valued for its dietary protein and calories, and favored in sustainable crop rotations for its ability to fix atmospheric nitrogen. Common bean production is often confronted with abiotic and biotic stresses including salinity and diseases like root rot commonly caused by *Fusarium* and *Rhizoctonia* species. Increased soil salinity affects nodulation and crop growth that can greatly inhibit yield. Currently, soil salinity is on the rise and clear interactions between salinity and pathogen, and its overall effect on common bean are unavailable. Methods used to evaluate its potential effect on fungal development included assessment of mycelial growth, fungal biomass, and spore germination in culture media amended with sodium chloride (NaCl). Preliminary results demonstrate overall reduction in mycelial growth, fungal biomass and spore germination with increasing NaCl concentrations in all the isolates tested, albeit with differences in the levels of tolerance. Additionally, expression of selected genes in common bean induced by combined salt and pathogen stresses using qPCR are in progress. This study will provide an insight into the potential direct effect of abiotic stress (salinity) on biotic stressors (pathogens) and simultaneous presence of abiotic and biotic stress on host gene expression.

P0806: Legumes, Soybean, Common Bean, and related

Understanding the Epigenomics of Nucleosome Occupancy in Common Bean during Drought Stress

Joey Cainong[1], Mayavan Subramani[1], Rita Hayford[2], Anusha Kalavacharla[3], Fedrica Williams[1], Vasudevan Ayyappan[1] and Venu (Kal) Kalavacharla[1], (1)Delaware State University, Dover, DE, (2)University of Delaware, Newark, DE, (3)Collegium Charter School, Exton, PA

Common bean (*Phaseolus vulgaris*), a protein-rich crop consumed worldwide, is severely affected by drought stress. Gene expression in response to drought stress is often influenced by chromatin structure. Nucleosome, the basic repeat unit of the chromatin, regulates gene expression and is an integral part of the transcriptome. Nucleosome positioning and occupancy influence gene expression in response to drought stress. It is therefore important to understand the *in vivo* location of the nucleosomes to assess its role in drought stress response. We investigated nucleosome occupancy in the common bean DREB6B gene using 6-week-old drought-stressed plants. Southern blot analysis of MNase-digested chromatin using DREB6B probe confirmed that the DREB6B coding region was bound by nucleosome(s). MNase-digested chromatin and ChIP-H3-pulled DNA showed less amplification in drought-stressed plants compared to non-drought plants. Hence, low nucleosome occupancy of the DREB6B coding region indicates enhanced gene accessibility for transcription factors to respond to drought stress in common bean. Our long-term goal is to understand genome-wide nucleosome positioning in crop plants in response to stress and the results presented in this work may serve as a means for understanding stress memory in common bean.

P0807: Legumes, Soybean, Common Bean, and related

Diversity of Guatemalan Climbing Bean (*Phaseolus vulgaris* L.) Germplasm

Maria Gabriela Tobar Piñón[1], Samira Mafi Moghaddam[1], Rian Lee[1], Julio Villatoro[2], Carlos Maldonano Mota[1], Luz de Maria Montejo Dominguez[1], Juan Osorno[1] and **Phillip McClean**[1], (1)Dept of Plant Sciences, North Dakota State University, Fargo, ND, (2)Instituto de Ciencia y Tecnología Agrícola (ICTA), Guatemala, Guatemala, Guatemala

The diversity of common bean (*Phaseolus vulgaris* L.), the most important food legume, is distributed from northern Mexico to northern Argentina. While bean diversity in Mexico and South America is highly characterized, studies of Central American germplasm are rare. In some countries in this region, such as Guatemala, beans are the principal protein source for peoples with some of the highest rates of chronic malnutrition and other dietary-related health problems in the world. These problems are mostly located in the Guatemalan highlands. A few studies, with limited population size, suggested Guatemalan climbing beans are another race in the Middle American gene pool, and that they may represent a source of new alleles for bean improvement in Guatemala and elsewhere in the world. In this study, supported by the USAID

Legume Innovation Lab, two Guatemalan climbing bean populations (n~1,100) were SNP genotyped using genotyping-by-sequencing and compared with other members of the Middle American and Andean gene pools. The diversity and population differentiation of the two populations were analyzed and results show that these Guatemalan climbing beans are strongly differentiated and equally diverse when compared to other common bean races. Association mapping identified previously unknown loci associated with altitudinal adaptation, seed shape, and bean rust and anthracnose resistance. These results demonstrate that Guatemalan climbing beans are a unique race and a potential source of new alleles for bean improvement programs.

P0808: Legumes, Soybean, Common Bean, and related

The International Mungbean Improvement Network – Mobilizing the Mungbean Genetic Diversity as a Source for New Traits for Crop Improvement

Roland Schafleitner[1], Ramakrishnan Madhavan Nair[2], Col Douglas[3], Tun Shwe[4], Mahbub AKM Alam[5], Aydita Pratap[6], Sanjeev Gupta[6] and Eric Huttner[7], (1)World Vegetable Center, Tainan, Taiwan, (2)The World Vegetable Center - South Asia, India, (3)Agri-Science Queensland, Hermitage Research Facility, Warwick, Queensland, Australia, (4)Myanmar Department of Agricultural Research, Yezin, Nay Pyi Taw, Myanmar, (5)Pulses Research Centre, Bangladesh Agricultural Research Institute (BARI), Gazipur, Bangladesh, (6)ICAR - Indian Institute of Pulses Research (IIPR), Kanpur, India, (7)ACIAR, BRUCE ACT, Australia

Mungbean (*Vigna radiata*) is an important food and cash crop in the rice based farming systems of South and Southeast Asia. Short crop duration, low input requirement, high global demand and the capacity to improve soils through nitrogen fixation make mungbean an ideal rotation crop for smallholder farmers. Repeated use of a limited number lines in mungbean variety improvement led to a narrow genetic base of the crop. Consequently current mungbean varieties lack key traits to cope with emerging pests and diseases. A minicore set of 296 accessions derived from the World Vegetable Center germplasm collection was produced to improve the access to new traits for breeders. The International Mungbean Improvement Network distributed the set collection to Australia, Bangladesh, India and Myanmar for multi-location field testing. The first trials showed large variation in phenology (days to 50% flowering and days to maturity), plant height, yield and quality parameters. In addition, accessions with tolerance to abiotic stresses such as heat and saline soils, and, most importantly, new sources for resistance to *Mungbean yellow mosaic disease* and dry root rot (*Macrophomina phaseolina*) were identified. The minicore set was genotyped with about 8,000 single nucleotide polymorphic markers. A genome-wide association study identified candidate loci for *Mungbean yellow mosaic* disease resistance on chromosomes 6 and 7 and on unmapped scaffold sequences. Validation of these loci and mapping of additional traits is ongoing to identify the genetic loci conditioning breeder-desired traits to facilitate crop improvement.

P0809: Legumes, Soybean, Common Bean, and related

Nitric Oxide-Mediated Proline Production and Redox Homeostasis Contribute to Methane-Induced Osmotic Stress Tolerance in Mung Bean

Yihua Zhang, College of Life Sciences,Nanjing Agricultural University, Nanjing, China

Previous reports discovered that methane (CH_4) has a beneficial effect on osmotic stress. However, corresponding signaling transduction is still not clear. In this work, polyethylene glycol (PEG) treatment for 48 h progressively stimulated the production of CH_4 in germinating mung bean seedlings. To mimic the physiological role of CH_4, we further discovered that exogenous CH_4 and sodium nitroprusside (SNP, a well-known nitric oxide releasing compound) not only triggered the production of nitric oxide (NO) in PEG-stressed mung bean, but also alleviate the inhibition of seed germination, and above responses could be obvious impaired by NO scavenger(s). CH_4-induced stress tolerance was dependent on NO-mediated proline generation. Reactive oxygen species (ROS) detection, histochemical staining and lipid peroxidation tests showed that CH_4 could reestablish redox balance. The application of tungstate (an inhibitor of the nitrate reductase, NR) and N^{ω}-nitro-$_L$-Arg methyl ester hydrochloride ($_L$-NAME, an inhibitor of NO synthase in mammalian, NOS), further suggested that NR and NOS-like protein might be partially involved in above CH_4-stimulated responses. *In vitro* and scavenger tests showed that NO-mediated *S*-nitrosylation might be associated with CH_4 responses. Together, CH_4-induced osmotic stress was partially dependent on NO-mediated proline generation and redox homeostasis.

P0810: Legumes, Soybean, Common Bean, and related

Gene Identification in Faba Bean – to Synteny and Beyond

Donal M. O'Sullivan[1], Deepti Angra[2], Vasiliki Tagkouli[2], Khalil Khamassi[3], Walid El-Rodeny[4] and Mahmoud Zeid[5], (1)University of Reading, Reading, United Kingdom, (2)University of Reading, READING, United Kingdom, (3)Institut National de la Recherche Agronomique (INRAT), Tunis, Tunisia, (4)ARC-Sakha Agricultural Research Station, Kafr El-Sheikh, Egypt, (5)University of Alexandria, Alexandria, Egypt

In this talk, I will outline progress in the further exploitation of the syntenic relationships between *Vicia faba* and *Medicago truncatula* established in earlier work in pursuit of genes underlying a series of faba bean traits of interest.

The first example is *Dwarf1* (*Dwf1*) gene, a gibberellic acid-sensitive dwarfing gene, where synteny was used to identify a candidate gene, whose causative role was further confirmed by fine-mapping, allele re-sequencing and metabolite analysis.

The second example is the *VC* gene controlling a 10-fold reduction in the anti-nutritional factors vicine and convicine. The trait maps in a segment of *Vf* chr 1 which shows strong colinearity with Mt chr 2. Since only a handful of spp in the genus *Vicia* make these secondary metabolites, it is highly unlikely that the Medicago genome contains a strict orthologue of the *VC* gene. Therefore, the role of synteny in this case is to saturate the interval with markers which is being used as a basis for the ongoing positional cloning of the gene.

Ultimately, whilst synteny has been a useful framework to translate knowledge of gene function from model to crop species, traits such as *VC* and tolerance to the parasitic weed *Orobanche* challenge us to move beyond a dependency on generating marker coverage and causal gene hypotheses exclusively based on synteny. With this in mind, we have recently developed a 50K Axiom SNP genotyping array and will present validation of its use in greatly narrowing the *vc* region and present plans for a major new initiative to map *Orobanche* tolerance.

P0811: Legumes, Soybean, Common Bean, and related

Association of Inflorescence Architecture Traits with Synchronous Pod Maturity in Mungbean

Eunsoo Lee, Seoul National University, Seoul, South Korea

Mungbean (*Vigna radiata* (L.) R. Wilczek) is a one of legume crops cultivated mostly in South, East and Southeast Asia with a short life-cycle and a self-pollinating ability. Both the production and consumption of mungbean have increased steadily around the world, but one of challenges interfering with an efficiency of harvesting is a non-synchronous pod maturity requiring more labor because of several harvests. In this study, we found an association of inflorescence architecture traits with synchrony in mungbean by investigating growth and developmental habits of inflorescence architecture. Typically, mungbean has a compound raceme inflorescence architecture consisting of a main branch (primary) and multiple secondary branches that can produce flowers. However, we found one exceptional genotype named 'Binh khe D.X.' that has a simple raceme inflorescence architecture where flowers are produced directly from a main (primary) branch. As a result of comparing two groups, synchrony and non-synchrony, the difference in synchronous and non-synchronous pod maturity was caused by the degree of indeterminate characters such as branches and peduncles. This study suggests a preferable standard for inflorescence architecture traits for future breeding, as well as for genetic research in mungbean pod maturity synchrony.

P0812: Legumes, Soybean, Common Bean, and related

Genomic Analysis of the Effects of Domestication on Runner Bean (*Phaseolus coccineus* L.)

Elena Bitocchi[1], Elisa Bellucci[1], Andrea Benazzo[2], Chunming Xu[3], Kyung Do Kim[3], M. Carolina Chavarro[3], Jin Hee Shin[3], Giorgio Bertorelle[2], Scott A. Jackson[3] and Roberto Papa[1], (1)Department of Agricultural, Food and Environmental Sciences, Polytechnic University of Marche, Ancona, Italy, (2)Università degli Studi di Ferrara, Ferrara, Italy, (3)University of Georgia, Athens, GA

The introduction of next generation sequencing technologies has revolutionised genetic analysis, creating new opportunities for plant research that can also be applied to non-model species. We used RNA sequencing (RNA-Seq) technology to investigate the whole runner bean (*Phaseolus coccineus* L.) transcriptome of 29 wild and domesticated accessions; we added further 19 accessions as controls, including wild and domesticated accessions of *P. dumosus, P. acutifolius, P. lunatus* crop species and P. *microcarpus, P. angustissimus* and *P. leptostachyus* wild species. The materials were chosen on the basis of previous molecular analyses carried out using both multilocus molecular markers and nucleotide data, as well as passport information in order to have a set of accessions representative of the genetic diversity of the species. We identified an high number of single nucleotide polymorphisms and we used this large dataset for population genetics inference. In particular, variable RNA fragments in wild and domesticated forms were analysed in order to obtain a picture of the effects of domestication process on runner bean genome and to identify genes/ genomic regions putatively under selection during domestication. This information is crucial not only to reconstruct the evolutionary history of this species, but also for a more efficient conservation of the existing germplasm and for selective breeding programmes.

P0813: Legumes, Soybean, Common Bean, and related

A Genome-Wide Association Study (GWAS) Reveals the Genetic Basis of the Domestication Syndrome in Lima Bean (*Phaseolus lunatus* L.)

Paola Hurtado[1], Jorge Duitama[2], Daniel Debouck[3] and María I. Chacón S.[1], (1)Universidad Nacional de Colombia, Bogotá, Colombia, (2)Universidad de los Andes, Bogotá, Colombia, (3)International Center for Tropical Agriculture – CIAT, Cali, Colombia

Artificial selection by humans during domestication has led to genetic and phenotypic changes, which together are known as the adaptive domestication syndrome. Knowledge of these genetic and phenotypic changes is essential to understand the current genetic structure and diversity of crops. *Phaseolus lunatus* is a domesticated species, which is cultivated in several regions of the world and shows wide adaptation to climates and soils. Here we provide results from the genotyping of 280 wild and domesticated accessions through GBS (Genotyping By Sequencing) and a phenotypic characterization of morphological traits related to the domestication syndrome, such as growth habit, seed weight, pod shattering, and phenological traits such as time to flowering and maturity. A GWAS analysis for these traits was conducted as well. Four SNPs were associated with seed weight located on chromosomes 2, 5, 7, and 8. Two SNPs were related to flowering time on chromosome 7. This is the first attempt to identify the genetic basis of the domestication syndrome in *Phaseolus lunatus* L.

P0814: Legumes, Soybean, Common Bean, and related

Developing Breeders' Tools: Identifying and Visualising Genotype to Phenotype Associations in the Common Pea and Its Wild Relatives

Kirstie Hetherington, Earlham Institute, Norwich, United Kingdom

Pea (*Pisum sativum*) is an important crop plant for food security. Cool season legumes are the second most important family of crops and are high in protein, making them important for food and animal fodder. Furthermore, peas are useful in crop rotation as they can symbiotically fix nitrogen, reducing the need for fertiliser.

The John Innes *Pisum* Collection is a well-characterised collection of peas, holding approximately 3,600 accessions. Previous work with retrotransposon-based insertion polymorphism markers (RBIP) characterised the collection into 3 distinct groups: cultivar, landrace and wild material (Jing *et al.,* 2010). We are undertaking a GWAS analysis of a number of important traits using this collection

We generated a core collection of representatives across all 3 groups consisting of 350 accessions. Biological replicates were sown in field and glasshouse environments for phenotypic and sequence analysis. To quantify phenotypic traits, we developed a novel automated image analysis tool, to measure simple and morphological shape descriptors (SMSDs) from images of pea leaflet, pod and seed. We also collected data on seed weight and plant height

The Pea genome is large and highly repetitive and currently has no reference. To tackle this issue, we have used Genotyping-by-sequencing (GbS) to generate a high-density marker panel of SNPs at low cost. Here we present an update on the GWAS analysis.

P0815: Legumes, Soybean, Common Bean, and related

Expanding Cowpea Genetic and Genomic Resources

Maria Muñoz-Amatriain[1], Stefano Lonardi[2], Sassoum Lo[1], Steve Wanamaker[1], Qihua Liang[2], Shengqiang Shu[3], Tingting Zhu[4], Ming-Cheng Luo[4] and Timothy J. Close[1], (1)Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA, (2)Department of Computer Science & Engineering, University of California, Riverside, CA, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)Department of Plant Sciences, University of California, Davis, Davis, CA

Cowpea, *Vigna unguiculata* L. Walp, is a diploid warm-season legume with a genome size of ~620 Mb. Cowpea is one of the most important crops for subsistence farmers in sub-Saharan Africa, and it is also widely grown in Latin America, parts of Asia, Southern Europe, and in the United States. Diverse cowpea germplasm is available from collections worldwide, the largest of which is held by the International Institute of Tropical Agriculture (IITA, Nigeria).

Much progress has been accomplished towards the development of genomic resources. This includes an iSelect genotyping array for over 50,000 SNPs, several genetic maps and a consensus map, and a reference genome sequence of the African variety "IT97K-499-35". The reference genome has been produced using PacBio sequencing (91x coverage) together with two BioNano optical maps and ten genetic maps containing 44,003 unique SNPs. The v1.0 cowpea pseudomolecules contains 519.4 Mb of superscaffold sequences with N50 = 16.4 Mb and L50 = 12, and it is now available on Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Vunguiculata_er). Sequencing of additional genomes is underway, including the popular Californian cultivar "CB46". In addition, significant effort has been made to genetically characterize collections of diverse cowpea germplasm to identify and incorporate beneficial alleles into breeding programs. An example of this is the "UCR Minicore", a set of 384 accessions assembled based on prior genetic and phenotypic knowledge that contains cultivated cowpeas from 51 countries on six continents. This minicore has been genotyped and phenotyped, and GWAS is resulting in many marker-trait associations. iSelect genotyping of the entire IITA Core Collection (2,032 accessions) is currently underway.

This work was supported by the NSF BREAD program and the Feed the Future Innovation Lab for Climate Resilient Cowpea.

P0816: Legumes, Soybean, Common Bean, and related

Genome-Wide Association Studies Identify Genomic Regions Controlling Seed Size Traits in Cowpea

Sassoum Lo[1], Maria Muñoz-Amatriain[1], Ndiaga Cisse[2], Philip A. Roberts[3] and Timothy J. Close[1], (1)Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA, (2)Institut Senegalais de la Recherches Agricole, Dakar, Senegal, (3)Department of Nematology, University of California Riverside, Riverside, CA

Cowpea (*Vigna unguiculata* L.Walp), is one of the most important grain legumes in sub-Saharan Africa, frequently used as the main source of protein in the human diet. Increasing seed size is a major breeding target trait as it influences cowpea's market price and is a determinant of grain yield. Understanding the genetic basis of seed size is fundamental to meeting future productivity requirements and to increasing market demand. In this study, genome-wide association studies (GWAS) were performed to dissect the genetic basis of seed size using a diverse minicore collection of 384 accessions. This minicore was genotyped using the Cowpea iSelect Consortium Array consisting of 51,128 SNPs. Seed size-related traits including 100-seed weight, seed length, and seed width were evaluated across three production locations. GWAS identified genomic regions that were significantly associated with seed size on chromosomes 3, 5,6,10 and 11. Based on annotations of the cowpea reference genome, candidate genes in these regions were noted. This study provides a basis for deciphering the genetic architecture of seed size in cowpea. This work was supported by the Feed the Future Innovation Lab for Climate Resilient Cowpea.

P0817: Legumes, Soybean, Common Bean, and related

Mapping Consumer-Related Seed Coat Traits in Cowpea (*Vigna unguiculata* L. Walp)

Ira A. Herniter[1], Sassoum Lo[2], Yi-Ning Guo[1], Maria Muñoz-Amatriain[2] and Timothy J. Close[2], (1)Department of Botany & Plant Sciences, University of California - Riverside, Riverside, CA, (2)Department of Botany & Plant Sciences, University of California Riverside, Riverside, CA

Seed coat color is among the most important consumer-related traits for cowpea (*Vigna unguiculata* l. walp). Cowpea is a warm season legume cultivated in sub-Saharan Africa, Southeast Asia, South America, the Mediterranean, and the United States. Cowpea is an especially important crop in sub-Saharan Africa, where it is a major source of protein. Consumers make qualitative judgements about products based on visual aspects, including color and pattern. Additionally, each market has particular preferences for seed coat color and pattern. As such, it behooves plant breeders to understand the genetic basis of these traits. This will allow newly developed varieties to be wrapped in the seed coat appropriate for a given market. Efforts to understand the basis of seed coat traits date back to the early twentieth century. With access to more highly advanced QTL mapping techniques and a high-throughput SNP genotyping system we have successfully mapped numerous seed coat traits, including seed coat texture and brown color, both highly relevant consumer-related traits, on chromosomes Vu08 and Vu03, respectively. Candidate genes underlying all the mapped traits will be identified and PCR markers developed for future application in breeding programs.

P0818: Legumes, Soybean, Common Bean, and related

Genomic Integrated Cowpea Breeding for Yield Potential and Stress Resilience in Uganda

Patrick O Ongom[1], **Isaac Onziga Dramadri**[1], Richard Edema[1], Kassim Sadik[2], Awio Bruno[1] and Paul T Gibson[1], (1)Makerere University, Kampala, Uganda, (2)Abi Zonal Agricultural Research and Development Institute (Abi ZARDI), Arua, Uganda

Cowpea *(Vigna unguiculata* L. Walp.) is self-pollinating, with one of the smallest genome sizes among legumes (2n = 2× = 22, ~620 Mbp). Cowpea is protein rich, and supplies grain, edible leaves and immature pods, and forage in the semi-arid tropics, mostly for substance farmers. Despite its importance, cowpea remains less exploited due to limited research attention, but interest has increased due to increasing global food and nutrition demand. Past crop genetic improvement was largely through phenotypic selection and recombination of existing natural variation. Genomics, through advances in molecular technologies and statistical capabilities, has begun to enhance plant breeding programs. Effective exploitation of this opportunity, especially in poorly-resourced breeding programs, will require innovative readjustment of breeding designs. We present here an integrated breeding approach aimed at enhancing development of high yield and stress tolerant cowpea varieties for Uganda. Starting with an initial germplasm base of 250 world collection, we have expanded this to include 384 cowpea mini core collections

and 260 MAGI lines from the University of California Riverside, that are fully SNP genotyped. Preliminary results from the phenotyping of these base germplasm revealed variation for key agronomic traits. We have exploited the SNP data and phenotype to chose parents for hybridization. The eventual lines, are being evaluated for yield potential and resistance to key diseases, virus and insects. The data is further being utilized to decipher the underlying genetics of key traits for eventual application in marker assisted selection. This integrated approach should enable more rapid and cost-efficient development of higher-yielding, stress-resilient, farmer and consumer-preferred cowpea varieties.

P0819: Legumes, Soybean, Common Bean, and related

Evaluation of Figs Agro-Biodiversity under Terminal Drought in Kabuli-Type *Cicer arietinum*

Kay Watt, Matisse Madrone, Antonia Palkovic and Paul Gepts, University of California, Davis, Davis, CA

Terminal drought is the major abiotic stressor in chickpea. It affects over 80% of the global production of chickpea and is the cause of more than 40% yield loss. Two studies were conducted to determine if general drought-associated traits were relevant to chickpea. Thirty-six chickpea varieties were selected using the Focused Identification of Germplasm Strategy (FIGS) from a combination of local northern California commercial lines, University of California, Davis-released lines, and varieties collected from areas with historically low average annual rainfall. Traits such as maturity dates, leaf carbon isotope ratios over life stages, canopy architecture, and multiple yield parameters were measured in greenhouse and field environments under terminal drought and control treatments. The range in drought seed mass between varieties was around 3kg / 120 plant plot in the field and 5g / plant in the greenhouse. Resilience under terminal drought was considered a minimal reduction in yield between the control and drought treatments. Yield was defined in several different ways, and several highly significant correlations were observed. The life stage significantly influenced the correlation of the observation to the yield data. As an example, water use in the vegetative stage and the reproductive stage were negatively and positively correlated with yield, respectively. The greenhouse and field trials allowed both an in-depth study of water use over time as well as observation of plant performance while competing for resources under terminal drought.

P0820: Legumes, Soybean, Common Bean, and related

Mapping of Yield Associated QTLs in Cultivated x Wild Populations of Chickpea

Sehrish Ijaz, UC Davis, Davis, CA and Noelia Carrasquilla-Garcia, Peter Chang and Douglas Cook

Chickpea (*Cicer arietinum*) is a self-pollinated, diploid grain legume crop, mainly cultivated in arid and semiarid regions of the world. It ranks third in worldwide production of food legumes and is an abundant source of highly nutritious and relatively inexpensive protein in developing countries. Its economic importance justifies the development and application of extensive genomic resources for yield enhancement. However, chickpea cultivated varieties have a very narrow genetic base, which hinders the improvement of numerous traits, including yield. In this regard, wild relatives of chickpea are a promising source of genetic and phenotypic variation for improving yield, as wild relatives have broad range of genomic diversity for a range of desirable agronomic features. In the present study, genotyping-by-sequencing (GBS) libraries were constructed for twenty-three segregating populations derived from a recurrent cultivated parent crossed with twenty-three diverse wild accessions. The wild accessions correspond to two sister *Cicer* species, *C. reticulatum* and *C. echinospermum*, collected from a range of natural populations and environments in south eastern Turkey. Sequencing was conducted by Illumina Hiseq 4000 and SNP calling was performed after mapping to a common cultivated reference genome. SNPs linked to yield traits, i.e. plant biomass, total seed biomass, pod shattering, plant volume, biomass conversion efficiency, and seed weight, were analyzed using a quantitative trait locus (QTL) mapping approach.

P0821: Legumes, Soybean, Common Bean, and related

Conservation Genetics of South Florida Tephrosia

Eric Bishopvonwettberg, University of Vermont, Burlington, VT

Several populations of *Tephrosia,* the hoarypeas (Fabaceae), exist in South Florida and the neighboring Caribbean Islands. To clarify relationships in this group, and to elucidate the conservation status of populations in Everglades National Park and Big Cypress Preserve, we used restriction site associated DNA sequencing (RAD-SEQ) on 94 samples from South Florida (Russel Key in Everglades, three stands in Big Cypress, Chapman Field, and Ludlam Pineland) and Puerto Rico (Sierra Bermeja, Cabo Rojo, and Conuco). Populations in Miami-Dade County and in Puerto Rico were collected with assistance by staff from Fairchild Tropical Botanic Garden. Restriction Site Associated DNA sequencing generated 6278 single nucleotide polymorphisms (SNPs). Analysis of variation in SNP markers by the Bayesian STRUCTURE algorithm and principle coordinate analysis both separated the samples into three groups, one representing the populations from Big Cypress, a second represented by Ludlum pineland, and the third both Russel Key, Chapman Field, and the three Puerto Rican populations. From these patterns we infer three taxonomic groups, with Big Cypress being a distinct taxa, putatively *T. seminole* or *T. curtissi* but also potentially a glabrous species that has not been well described from the Caribbean, Ludlum being *T. floridiana*, and Russel Key, Chapman Field, and Puerto Rico all belong to *T. angustissima*. Diversity is moderate in all taxa, with only limited evidence of a bottleneck in some of the disjunct populations such as Russel Key and Chapman Field.

P0822: Legumes, Soybean, Common Bean, and related

Genetic Diversity, Population Structure and Botanical Varieties of 320 Peanut Cultivars Revealed by Tunable Genotyping by Sequencing (tGBS)

Zheng Zheng, Industrial Crops Research Institute, Henan Academy of Agricultural Sciences, Zhengzhou, China

To reveal the genetic evolutionary relationship of different botanical varieties of *Arachis hypogea* L., we collected 320 peanut accessions around the world, which consisted of landrace and released cultivars. A total of 3718 high quality single nucleotide polymorphisms (SNP) were filtered from 1,240,787 polymorphic sites using tunable Genotyping by Sequencing (tGBS). Phylogenetic tree indicated that 320 peanut accessions were clustered into three groups: almost all peanut accessions in the first group were var. *fastigiata*, the second group were consisted of var. *vulgaris* from China and runner type from other countries and the third group included the intermediate varieties and var. *hypogea*. The results of population structure and PCA analysis were consistent with that of phylogenetic tree. Fixation index (F_{ST}) value and sequence diversity (π) ratios indicated that the var. *fastigiata* was near to var. *vulgaris* (F_{ST}= 0.2838) and var. *hypogea* was significantly distinct from the

other two botanical varieties (F_{ST}>0.4). Moreover, 97 homozygous and polymorphic KASP were developed and verified, which can be used in marker-assisted breeding and peanut cultivar identification.

P0823: Legumes, Soybean, Common Bean, and related

The SNP-Based High-Density Linkage Maps and Fine Mapping of QTLs Controlling Disease Resistance using Whole-Genome Re-Sequencing in Peanut

Baozhu Guo[1], Gaurav Agarwal[2], Josh Clevenger[3,4], Manish K Pandey[5], Hui Wang[2], Yaduru Sasidar[5], Sunil Gangurde[5], Ye Chu[2], Jake Fountain[2], Divya Choudhary[2], Albert Culbreath[2], Xin Liu[6], Rupesh Deshmukh[7], Corley Holbrook[8], David J. Bertioli[3], Peggy Ozias-Akins[2], Scott A. Jackson[3] and Rajeev Varshney[5], (1)USDA-ARS, Crop Protection and Management Research Unit, Tifton, GA, (2)University of Georgia, Tifton, GA, (3)University of Georgia, Athens, GA, (4)Mars Wrigley Confectionery, Athens, GA, (5)ICRISAT, Hyderabad, India, (6)BGI-Shenzhen, Shenzhen, China, (7)University Laval, Quebec City, QC, Canada, (8)USDA-ARS, Crop Genetics and Breeding Research Unit, Tifton, GA

High-density genetic linkage maps are essential for fine mapping QTLs controlling disease resistance traits, such as early and late leaf spots (ELS, LLS), Tomato spotted wilt virus (TSWV). With completion of the genome sequences of two diploid ancestors of cultivated peanut, we could use whole-genome re-sequencing (WGRS) technology to genotype recombinant inbred line (RIL) populations and develop high-density genetic maps for peanut. We constructed the first sequence-based high density maps with a total of 8,869 and 14,500 SNPs assigned to 20 linkage groups, representing the 20 chromosomes, for the "T"- and the "S" populations, respectively. The total length of the linkage maps were 3,120 and 3,201 cM with an average distance of 1.45 and 0.93 cM among 2,156 and 3,400 loci for the "T" and the "S", respectively. The genetic maps showed both homoeologous and translocated markers with the "T" having 739 as homeologus and 413 as translocated markers, while the "S" showed 2422 SNPs as homeologus and 852 as translocated markers. For the "T"-population, there were a total of 35 main-effect QTLs (M-QTLs) for all three diseases with phenotypic variation explained (PVE) ranging from 6.32 to 47.63%. QTL with above 40% PVE were detected for each of the three diseases. QTL analysis revealed that a segment of chromosome A03 features major QTLs for ELS, LLS and TSWV. KASP markers were developed and validated for the SNPs associated with major QTLs, which could be used in genomics-assisted breeding (GAB).

P0824: Legumes, Soybean, Common Bean, and related

Fatty Acid Profiling of Four Different Peanut FAD2 Genotypes at Five Seed Development Stages

Ming Li Wang[1], Charles Y. Chen[2], Brandon Tonnis[1], David Pinnow[1] and Phat Dang[3], (1)USDA-ARS, PGRCU, Griffin, GA, (2)Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, (3)National Peanut Research Laboratory, Dawson, GA

Peanut is one of the most important edible oilseed crops. The level of oleic acid in peanut seeds can significantly affect the oil quality. Consuming peanut product from high oleic acid seeds may positively contribute to improving human health. The level of oleic acid in peanut seeds is mainly controlled by two pairs of homoeologous genes (*FAD2A* and *FAD2B*). Eight high-generation breeding lines were developed by selection from peanut breeding programs and their genotypes were determined and classified by real-time PCR and sequencing into homozygous *AABB*, *aaBB*, *AAbb*, and *aabb* (two of each genotype). These eight lines were grown at Dawson, GA in two replicates for two years. Fresh seeds were collected from five pod development and seed maturity stages (yellow 1, yellow 2, orange, brown, and black). After drying, the seeds were used for seed composition and fatty acid analysis. Our results showed: (1) four genotypes do not significantly affect protein and oil content but do significantly affect fatty acid profile; (2) as peanut pod development and seed maturation (from yellow 1 to black stage) progress, protein content increases but not significantly whereas seed weight and oil content increase significantly; (3) the level of oleic acid (C18:1) significantly increases whereas the levels of linoleic acid (C18:2), eicosenoic acid (C20:1), behenic acid (C22:0), and lignoceric acid (C24:0) significantly decrease; and (4) the levels of palmitic acid (C16:0), arachidic acid (C20:0), and cerotic acid (C26:0) generally stay unchanged. The information provided here would be useful for peanut breeders, product processors, and consumers.

P0825: Legumes, Soybean, Common Bean, and related

Genome-Scale Characterization of Wild Peanut Species of Section Arachis

Alicia N. Massa[1], Renee S. Arias[1], Shyamalrau P. Tallury[2], Victor S. Sobolev[1], Ronal B. Sorensen[1], Marshall C. Lamb[1] and Thomas H. Stalker[3], (1)USDA-ARS National Peanut Research Laboratory, Dawson, GA, (2)USDA-ARS Plant Genetic Resources Conservation Unit, Griffin, GA, (3)North Carolina State University, Raleigh, NC

Wild *Arachis* species are valuable genetic resources for peanut improvement. Among the eighty species that have been described in the genus, only taxa in section *Arachis* hybridize with cultivated peanut (*Arachis hypogaea*). In this study, a total of 168 accessions of 27 species of section *Arachis* were genotyped using a high throughput single nucleotide polymorphism (SNP) array. A subset of 150 accessions were evaluated for resistance to early leaf spot (*Cercospora arachidicola*) and late leaf spot (*Cercosporidium personatum*). Patterns of genetic variation within and among species were resolved with more than seven thousand high-confidence SNPs distributed across the ten peanut chromosomes. The presence of both early and late leaf spot in the field enabled the selection of germplasm against the diseases. Collectively, this study provides the molecular basis for identification and genomic introgression of novel variation into cultivated peanut.

P0826: Legumes, Soybean, Common Bean, and related

Development of a SNP-Based Map of a Peanut Wild Species Introgression Population by High-Throughput Analysis

Theophilus Kwabla Tengey[1], Ratan Chopra[2], Charles Simpson[3], Andrew Hillhouse[4], Venugopal Mendu[1] and **Mark D. Burow**[5], (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)USDA- ARS, Lubbock, TX, (3)Texas AgriLife Research, Stephenville, TX, (4)Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine, College Station, TX, (5)Texas A&M AgriLife Research, Lubbock, and Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

A genetic linkage map for a BC_1 interspecific peanut introgression population has been developed from SNP-based markers. A-genome and B-genome SNPs were selected from a transcriptome sequence database made of A-genome parents, A. *diogoi*, A. *cardenasii* and A. *duranensis*, and B/K-genome parents, A. *ipaënsis* and A. *batizocoi*, and parents of the backcross population (Florunner and TxAG-6). KASP primers synthesized were validated against parents using the Roche LightCycler. Validation results showed that 92 out of 124 selected A-genome SNPs (74.2%) and 128 out of the 178 selected B-genome SNPs (71.9%) perfectly matched the selection criteria used in targeting the SNPs. Markers were then scored on 64 BC_1 individuals, and a genetic map was developed, to date consisting of 150 SNP markers on 20 linkage groups. Thereafter, markers were scored on 317 individuals of a BC_3F_6 population on a Fluidigm Biomark HD. Markers are being used for identification of QTLs for resistance to leaf spots.

P0827: Legumes, Soybean, Common Bean, and related

Validation of New Affymetrix SNP-Array for QTL Mapping of Seed Composition Traits in Virginia-Type Peanut

Abhinandan Patil, ARO, Volcani Center, Rishon LeZion, Israel, Ilan Hedvat, ARO, Rishon LeZion, Israel, Peggy Ozias-Akins, University of Georgia Tifton Campus, Tifton, GA, Ye Chu, UGA, Tifton, GA, Josh Clevenger, University of Georgia, Athens, GA and **Ran Hovav**, ARO - Agricultiral Research Organization (Volcani), Bet Dagan, Israel

Improving seed quality traits is one of the most important goals in peanut breeding. Since many of these traits are relatively difficult to phenotype, an alternative indirect selection by markers is needed. Yet, peanut has very low degree of polymorphism especially among cultivated varieties that hindered the development of mapping tools. The aim of this study was to validate the usage of a new Affymetrix SNP-array as a tool for QTL mapping in cultivated peanut, with emphasis on seed composition traits. Two hundred and fifty-four RILs derived from the hybridization between two closely related Virginia-type varieties were used. Each line was phenotyped in three environments for percent oil, protein and total sugar content. Significant and high negative correlation was found between the lines in oil and protein content (-0.76) with moderate broad-sense heritability estimates. Genomic DNA was micro-arrayed with 2882 poly-high-resolution SNPs after filtering from total of 47,837 SNPs on the SNP-array. To initially inspect the robustness of the analysis the monogenic plant architecture trait with known genomic location was used. Ten SNP markers significantly co-segregated with the plant architecture phenotype, all are located directly in the known locus (B5 ~146M). Four significant QTL were found for both oil and protein content in the same loci, with the highest on Ch.A6 that explained 28% of the total phenotype, indicating that same genomic factors control the protein and oil content in this background. This study demonstrates the straight-forward utilization of DNA SNP-array for trait mapping into the low polymorphic commercialized peanut.

P0828: Legumes, Soybean, Common Bean, and related

Assessing Genetic Drift in the Alfalfa Cultivar Cuf 101 Using Genotyping-By-Sequencing

Scott Newell, UC Davis, Davis, CA and E. Charles Brummer, University of California, Davis, Davis, CA

Alfalfa cultivars are synthetic populations. Initial breeder seed is used to produce foundation seed, which is then grown to produce the certified seed class that is sold commercially. Initial breeder and foundation seed stocks will be expended if a cultivar has a long lifespan. Consequently, new foundation seed can be multiplied from an earlier generation of foundation seed. We hypothesized that this repeated seed increase could lead to genetic shifts due to genetic drift or natural selection. To examine the extent of genetic change over generations of foundation seed increase, we evaluated the popular and long standing non-dormant alfalfa cultivar CUF-101. Since its release more than 40 years ago, CUF-101 has required ten independent foundation seed increases. To assess whether genetic drift or natural selection has occurred in CUF-101, we genotyped the ten foundation seed lots together with several seed lots collected from certified seed fields currently in production in the Imperial Valley, CA. Using bulked DNA samples of ~100 individuals from each population, Genotyping-by-Sequencing (GBS), and the GBS-SNP-CROP pipeline, we identified allele frequencies in each population at loci throughout the genome. We then used principal components analysis to assess the genetic similarity of the different foundation and certified seed populations. Our results showed that there has been very little drift across CUF-101 foundation seed lots over time. However, we noted differences between these foundation seed lots and some certified seed lots, which necessitates further investigation.

P0829: Legumes, Soybean, Common Bean, and related

Hydrogen-Induced Osmotic Tolerance is Associated with Nitric Oxide-Mediated Proline Accumulation and Reestablishment of Redox Balance in Alfalfa Seedlings

Jiuchang Su, Nanjing Agricultural University, Nanjing, China

Although hydrogen (H_2) and nitric oxide (NO) are respectively suggested to enhance plant tolerance against osmotic stress, the corresponding causal link is still elusive. In this report, the application of hydrogen-rich water (HRW) strengthened the production of NO in PEG-stressed alfalfa seedling roots, followed by the obvious alleviation of seedling growth inhibition. Comparatively, significant but weaker responses in phenotypes were observed in the plants supplemented with nitrogen-rich water, indicating that the role of HRW was H_2-related. Above responses of H_2 were inhibited by the removal of NO with the scavenger(s) of NO. The application of tungstate, an inhibitor of the NO synthetic enzyme nitrate reductase (NR), showed the similar blocking response in the phenotype, suggesting that NR might be the major source of NO involved in above H_2 actions. Proline synthesis was also stimulated by H_2 and NO, both of which were supported by the increased Δ^1-pyrroline-5-carboxylate synthetase (P5CS) activities, the decreased proline dehydrogenase (ProDH) activities, and corresponding transcripts. The addition of H_2 and NO could increase antioxidant defence in stressed plants, both of which were confirmed by the histochemical staining for reactive oxygen species (ROS) production and lipid peroxidation, representative antioxidant enzyme activities and transcripts. Thus, redox balance was reestablished. When NO scavenger was applied, proline synthesis, redox balance, and thereafter osmotic tolerance induced by H_2, were severely impaired. Additionally, H_2-triggered *S*-nitrosylation was obviously inhibited by the removal of endogenous NO level. Together, above results discovered the involvement of NO-induced proline and redox balance in H_2-triggered osmotic tolerance.

P0830: Legumes, Soybean, Common Bean, and related

Genomic Architecture and Phenotypic Plasticity of Forage Quality in Response to Water Deficit in Alfalfa (*Medicago sativa* L.)

Long-Xi Yu[1], Bill Boge[1], Charles Hawkins[1] and Steve Fransen[2], (1)USDA-ARS Plant Germplasm Introduction and Testing Research, PROSSER, WA, (2)Irrigated Agriculture Research and Extension Center, Washington State University, PROSSER, WA

Plant phenotypic plasticity is the ability for plants to cope with environmental factor variability. However, mechanisms by which phenotypic plasticity affects plant adaptation to environmental change remain largely unknown. It is important to identify plant functional traits in which plasticity may play a critical role in plant response to the environmental change. In the present study, we characterized 31 forage quality traits in alfalfa populations and analyzed the phenotypic plasticity of these traits in response to a gradient of water deficit. The plasticity index (PI) varied among the traits with the highest PI value (1.1) for K content and lowest (0.2) for dry matter. Fiber contents such as lignin, ADF and NDF decreased as drought increased. In contrast, energy traits were increased as drought increased. Correlation coefficients between the quality traits were increased as drought increased. Genetic factors were also characterized using genotyping by sequencing and genome-wide association studies. Single nucleotide polymorphisms associated with the traits were identified. Genomic architectures for phenotypic plasticity were analyzed for each trait and compared between the traits. Genomic regions responsible for the traits were identified and characterized. The possible roles of the genetic factors affecting phenotypic plasticity of forage quality in alfalfa were discussed.

P0831: Legumes, Soybean, Common Bean, and related

Network Assisted Analysis of *Medicago truncatula* Transcriptional and Metabolic Interactions

Rens Holmer, Sandra Smit, Rene Geurts and Wouter Kohlen, Wageningen University, Wageningen, Netherlands

As a model system for nitrogen fixing plant-microbe symbiosis, the legume *Medicago truncatula* has been studied by numerous research groups. Over two decades of research have resulted in key insights in the role of transcriptional and hormonal regulation in the symbiosis. A plethora of recent RNA sequencing efforts have resulted in ~500 available transcriptome samples in public repositories. Whereas the individual experiments for which these samples have been generated all have provided valuable insights in the symbiosis, genome-wide analysis of the interaction between transcriptional and hormonal regulation is currently lacking. Since studying these interactions will be crucial to further our understanding of the symbiosis, a system that can integrate gene-centric and metabolite-centric views and analyses is indispensable. Here we present an integrated framework for the analysis of *M. truncatula* transcriptome data and metabolic pathway information. We take advantage of public transcriptome data for gene coexpression analysis and leverage recent developments in graph database technology to create a system that can store, query and visualize co-expression data, individual experiments and metabolic pathways. We use this system in a combined transcriptome analysis of the *Mtcre1* cytokinin receptor and *Mtnin* transcription factor mutants. Based on this analysis we present new leads for studying the interaction between transcriptional and hormonal regulation in the legume-rhizobium symbiosis.

P0832: Legumes, Soybean, Common Bean, and related

Detecting Gene Expression Patterns Governing Root Nodule Development in *M. truncatula*

William L. Poehlman, Clemson University, Department of Genetics & Biochemistry, Clemson, SC

Root nodulation enables atmospheric nitrogen fixation in *Medicago truncatula* through rhizobial symbiosis. While this process is beneficial for both organisms, the number of nodules produced by the plant must be regulated in order to balance consumption of carbon with nitrogen. The process of initiating and regulating root nodule formation involves signaling pathways that are mediated by tissue-specific gene expression patterns over time. To elucidate these expression patterns, we have performed RNA sequencing (RNAseq) of samples extracted from the elongation zone of *M. truncatula* roots across five distinct time points, comparing plants inoculated with *Sinrhizobium meliloti* to uninoculated plants. Using differential gene expression and time-series analysis, we have identified RNA transcripts that are differentially expressed at specific time points following inoculation. In addition, we have constructed a gene coexpression network (GCN) using 232 public RNAseq datasets. We have found groups of differentially expressed transcripts from our RNAseq data that significantly overlap with GCN modules that were detected from this public data. Exploring the function of these differentially expressed transcripts in the network biology context will provide insight into the pathways that are necessary for root nodule formation and regulation, and provide novel candidate genes for functional validation.

P0833: Legumes, Soybean, Common Bean, and related

From the Freezer to the World: The Genomes of Allopolyploid White Clover and its Progenitors.

Andrew G. Griffiths[1], Roger Moraga[1], Stig Uggerhøj Andersen[2], Vikas Gupta[2], Timothy Bilton[3], Marni Tausen[4], Matthew A. Campbell[5], Rachael Ashby[3], Istvan Nagy[6], Anar Khan[7], Craig Anderson[1], Benjamin K. Franzmayr[1], Kerry Hancock[1], Alicia Scott[1], Nick W. Ellison[1], Murray P. Cox[8], Torben Asp[9], Thomas Mailund[10] and Mikkel Heide Schierup[2], (1)AgResearch, Palmerston North, New Zealand, (2)Aarhus University, Aarhus, Denmark, (3)AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand, (4)Department of Molecular Biology and Genetics, Aarhus, Denmark, (5)Department of Ecology and Evolutionary Biology, Santa Cruz, CA, (6)Aarhus University, Slagelse, Denmark, (7)AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand, (8)Massey University, Palmerston North, New Zealand, (9)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (10)Bioinformatics Research Centre, Aarhus, Denmark

The merging of distinct genomes, allopolyploidisation, generates adaptive potential through increased genetic diversity and access to ‘genomic toolboxes’ from the contributing genomes. White clover (*Trifolium repens* L.) is an allotetraploid (2n=4x=16) forage crop found throughout temperate grasslands, and is derived from two diploid progenitors: *T. occidentale* and *T. pallescens*, each confined to markedly different coastal and montane niches, respectively.

Genome and transcriptome sequencing and subsequent assembly of this species complex, has provided a wealth of data to gain insight into the genesis and evolution of white clover. We have confirmed the progenitors, and shown that the progenitor subgenomes within white clover have largely retained their integrity and gene expression activity following allopolyploidisation. Furthermore, we show that this hybridisation event occurred during the depths of the last glaciation at a time when the European progenitor ranges (coastal and montane) likely overlapped.

Born of climate change, white clover, therefore, represents a clear example of allopolyploidy-facilitated niche expansion, where the two progenitor genomes reunited and expanded from disparate and highly specialised European habitats to a ubiquitous global presence. Perhaps

underpinning this evolutionary success, we found high polymorphism levels in white clover, demonstrating diversity carry-over from its progenitors. Furthermore, we have also found evidence of tissue-specific expression switching between subgenome copies of genes involved in flavonoid biosynthesis, a key pathway involved in adaptive traits such as plant/microbial interactions.

P0834: Legumes, Soybean, Common Bean, and related

The Identification of Population-Specific Patterns in Breeding Programmes of *Trifolium repens* and Related Species

Lucy Marie Egan, AgResearch and Lincoln University, Christchurch, New Zealand, Valerio Hoyos-Villegas, AgResearch, Christchurch, New Zealand and Rainer Hofmann, Lincoln University, Christchurch, New Zealand

The *Trifolium* family is a key legume family in agroecosystems of the world. White clover (*Trifolium repens*) is the most common pasture legume in New Zealand where it is usually grown with perennial ryegrass in swards and grazed *in situ*. White clover breeding in New Zealand started in the early 1930's and due to the large economic value of white clover, there have been considerable research activities and development. The objective of this study was to identify population-specific patterns in *Trifolium* breeding programmes and to produce an overview of these patterns across time and human based decision making. Parental data for *T. repens* as well as other *Trifolium* species such as *T. arvense, T. ambiguum, T. dubium, T. hybridum, T. medium, T. pratense and T. subterraneum*, as well as an interspecific hybrid of *T. repens* x *T.* occidentale were collated from 1941-2016. These data sets, including breeding methods of bi-parental and single crosses and selections, backcrosses and polycrosses were constructed into pedigrees for the purposes of this study. Coefficients of coancestry and inbreeding were derived from biparental and single crosses in families. *T. repens* and *T. pratense* were the only species in the *Trifolium* family to show inbreeding. *T. pratense* had an average population inbreeding coefficient of 0.56% and a coancestry coefficient of 0.61%. Both commonly used families and unused families were identified from the pedigrees.

P0835: Legumes, Soybean, Common Bean, and related

Using DNA Barcoding to Re-Evaluate Clover Accessions in Germplasm Repositories.

Howard M. Laten, Christina S. Jreisat, Mohammad R. Salah, Ereni Kourkouvis and Mohammed A. Abourahma, Loyola University Chicago, Chicago, IL

Accurate identification and cataloging of plant genomic resources are essential for agronomy management, genetic diversity characterizations, and phylogenetic analyses. Expert collectors and seed bank staff and curators do their best to attain high levels of accuracy and reliability, but errors are inevitable. At least 16 clover species are actively cultivated worldwide and synthetic hybrids with enhanced growth characteristics and greater environmental tolerances have been sought for many years. An extensive phylogenetic study of the genus *Trifolium*, based on the DNA sequences of two "barcode" regions (nuclear ribosomal DNA internal transcribed spacer (ITS) and chloroplast *trnL* intron sequences) from 218 species, was published in 2006. Inter-species variation in these regions was sufficient to fully resolve closely related species and subspecies in this and in a second, smaller 2006 study. By amplifying and sequencing the ITS of clover accessions not included in the 2006 studies, we found and report here inconsistencies in the species designations of several accessions within and between major germplasm centers. Our goal is to provide the centers with candidate accessions for further investigation to identify the nature and sources of these inconsistencies. Although we are using a low throughput protocol, the project is ideally suited for an undergraduate research environment with modest resources.

P0836: Legumes, Soybean, Common Bean, and related

New Phase of the National Bioresource Project *Lotus* and *Glycine*

Shusei Sato[1], Masatsugu Hashiguchi[2], Hidenori Tanaka[2] and Ryo Akashi[2], (1)Graduate School of Life Sciences, Tohoku University, Sendai, Japan, (2)University of Miyazaki, Miyazaki, Japan

The National BioResource Project (NBRP) was launched by the Japanese government in 2002 with the objective of collecting, conserving and distributing such valuable, independent resources and making them easily available for the larger research community. The NBRP project has entered into 4th phase from April 2017, and the program of *Lotus japonicus* and *Glycine max* renewed its web database, "LegumeBase" (https://www.legumebase.brc.miyazaki-u.ac.jp/), as a kickoff of the new phase.

In the phase 4 of NBRP *Lotus* and *Glycine*, we are continuing our efforts on providing the material resources, such as seeds of experimental strains, wild accessions and recombinant inbred lines of *L. japonicus*, wild accessions of *G. soja*, recombinant inbred lines of *G. max*, full-length cDNA clones of *L. japonicus* and *G. max*, and signature tagged mutant lines of *Mesorhizobium loti*, a symbiont of *L. japonicus*. In addition, we are going to provide the information resources, such as updated reference genome sequence of *L. japonicus* accession Gifu, and genome resequence-based genotype information of RILs and wild accessions of *L. japonicus,* both of which were prepared with the support of NBRP Genome Information Upgrading Program. Also we are going to improve our material resources by increasing the number of native retrotransposon (*LORE1*) insertion tag lines of *L. japonicus*, collecting the published symbiotic mutant lines of *L. japonicus* and *G. max,* and establishing the collection of the pairs of wild accession of *L. japonicus* and its natural symbionts. These updated resources and information are available through "LegumeBase".

P0837: Legumes, Soybean, Common Bean, and related

Development of Genomic Resources and Genetic Markers for the Andean Lupin (*Lupinus mutabilis* Sweet)

Julio Solis-Sarmiento[1], Raul Blas-Sevillano[1], Jeny Perez-Huamanlazo[1], Dmitry A. Filatov[2] and Felix Camarena-Mayta[1], (1)UNIVERSIDAD NACIONAL AGRARIA LA MOLINA, Lima, Peru, (2)University of Oxford, Oxford, United Kingdom

Lupinus mutabilis Sweet (Andean lupin) is a legume of great importance to the highlands due to its high protein and oil content. However, its orphan condition has limited the generation of genomic tools to aid breeding efforts. The recent publication of the *L. angustifolius* L. (narrow-leafed lupin) genome has provided information to facilitate the discovery and characterization of important genes and markers in the genus *Lupinus*. In this study, we performed RNA-seq analysis of *L. mutabilis* and other lupin species from the Andean region. Annotation of transcripts were conducted by comparison against a databases of model organisms including the narrow-leafed lupin. Of the conserved functional markers found among lupin species, fifty simple sequence repeat (SSR) markers were screened for polymorphism using eight

accessions. The polymorphic SSRs were used to assess the diversity of forty-six accessions. These markers are highly informative for detecting genetic variations in Andean lupin. Additional work is being conducted to develop a draft genome of Andean lupin using the Illumina platform and progress made to this end are presented in this study. By mining both the transcriptome and genome data, we have found several pathways and genes of agronomic importance for the Andean lupin. The resources generated from the current investigation are important for the establishment of a modern lupin breeding program and sustainable germplasm management in the Andes.

P0838: Maize, Sorghum, Millet, Sugar Cane, and related

Germplasm Enhancement of Maize (GEM) - 24 Years of Public-Private Sector Collaboration to Increase US Maize Genetic Diversity

Candice Gardner, USDA-ARS and Iowa State University, Ames, IA, David Peters, USDA-ARS and Iowa State University and Matthew Krakowsky, USDA-ARS and North Carolina State University

The Germplasm Enhancement of Maize (GEM) Project is a mission-oriented, cooperative research effort of the USDA,ARS, land grant universities, private industry, and international agricultural research centers to broaden the germplasm base of maize cultivated within the US. While exotic germplasm can be defined as any germplasm that has not been sampled in a breeding program, the GEM project uses the term to cover landraces and improved germplasm from tropical and subtropical origins. The Raleigh location of the GEM Project is focused on identifying new exotic sources of maize germplasm and on developing 50% exotic/50% temperate germplasm, while the Ames location focuses on developing 25% exotic/75% temperate germplasm with high yield potential and resistance to common foliar, stalk and ear diseases which can be incorporated directly into commercial maize breeding programs. The environmental conditions and geographical latitude of the Raleigh location make it feasible to work with breeding material that contains a higher percentage of exotic germplasm than is usually practical at Midwestern locations such as Ames, IA, and this should enable the transfer of more genetic diversity into commercial programs. Private sector collaborators provide proprietary germplasm and in kind support in the form of trials, nursery resources, and various trait screening resources. University collaborators focus on specific traits or areas of interest, and have released improved germplasm for biotic and abiotic stress resistance, for novel starch properties, and to understand the nature of diversity. Graduate students in US public universities have gained experience with introgression breeding and maize genetic resources, critical training to provide for continued crop improvement. Currently there are 21 US public, 30 US private, 5 international public, and 7 international private sector, active program participants.

In addition to developing germplasm for use in maize breeding programs, the GEM project also develops new resources that will allow the introgression of useful alleles from agronomically inferior exotic sources. This is known as the Allelic Diversity (AD) project, and involves crossing and backcrossing accessions from all of the races of maize to formerly proprietary temperate inbreds to develop a panel of lines that represent the diversity of maize. The temperate-adapted AD double haploid lines can be used to screen for alleles of interest that would otherwise be unavailable to maize researchers. To date, GEM participants have released about 300 conventionally derived, diverse lines. Additionally, more than 200 double haploid lines have been released as part of the AD project. These resources are freely available without restriction for breeding and other research use and are viewable at http://www.public.iastate.edu/~usda-gem/ .

P0839: Maize, Sorghum, Millet, Sugar Cane, and related

MaizeGDB: New Resources for Maize Researchers

John Portwood[1], Ethalinda Cannon[2], David Schott[2], Kyoung Tak Cho[2], Martynas Snarskis[2], Miranda Dietze[2], Jesse R. Walsh[3], Lisa Harper[4], Margaret Woodhouse[5], Jack Gardiner[6], Mary Schaeffer[7] and Carson M Andorf[1], (1)USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, (2)Iowa State University, Ames, IA, (3)USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA, (4)USDA-ARS Corn Insects and Crop Genetics Research, Ames, IA, (5)ISU, Ames, IA, (6)Division of Animal Sciences, University of Missouri, Columbia, MO, (7)USDA ARS and University of Missouri, Columbia, MO

MaizeGDB, the USDA-ARS maize genetics and genomics database, is a highly curated, community-oriented informatics service to researchers focused on the crop plant and model organism *Zea mays*. MaizeGDB facilitates maize research by curating and maintaining a database that serves as the central repository for the maize community. With the availability of more reference quality genomes for maize, MaizeGDB has become more sequence-centric, while still maintaining traditional maize genetics datasets. The research focus of the maize community has continued to evolve, making it necessary to continually redefine data access and data analysis tools. In this poster we present an overview of new services and data types provided by MaizeGDB. New genome sequences are incorporated into MaizeGDB and made accessible through the annotation/assembly pages, BLAST databases, and genome browsers. Recently added genomes include B73v4, W22v2, Mo17, PH207, CML247, B104, EP1, F7, Ki11, and NC350. MaizeGDB is responsible for stewardship of the maize representative genome assembly (B73), including the improvement of associations between the B73 gene models and gene models for all other assemblies. New resources include CornCyc 8.0, a tool allowing users to query metabolic pathways on the B73v4 assembly and SNPversity, a tool allowing users to compare SNPs across a diverse set of inbred lines. New tools under development include MaizeMine (an InterMine instance), MaizeDIG (a tool for tagging phenotypes in images and linking them to genes), and PedNet (a tool for visualizing pedigree networks).

P0840: Maize, Sorghum, Millet, Sugar Cane, and related

MaizeGDB: Stewardship for Maize Genome Assemblies and Annotation

Ethalinda Cannon[1], Margaret Woodhouse[2], Lisa Harper[3], Jack Gardiner[4], John Portwood[5], Mary Schaeffer[6], Jesse R. Walsh[7] and Carson M Andorf[5], (1)Iowa State University, Ames, IA, (2)ISU, Ames, IA, (3)USDA-ARS Corn Insects and Crop Genetics Research, Ames, IA, (4)Division of Animal Sciences, University of Missouri, Columbia, MO, (5)USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, (6)USDA ARS and University of Missouri, Columbia, MO, (7)USDA-ARS, Corn Insects and Crop Genetics Research Unit, Ames, IA

MaizeGDB is the genetics and genomics database for the model organism and agriculturally important crop *Zea mays*. One of the main priorities at MaizeGDB is to provide genome assembly and annotation stewardship for the maize research community. With falling sequencing costs and improved genome assembly methods, it has become feasible to generate dozens of reference-quality genome assemblies for maize

accessions of importance to maize breeders and researchers. MaizeGDB currently hosts information for 10 high-quality genome assemblies (B104, B73, CML247, EP1, F7, Ki11, Mo17, NC350, PH207, and W22) and has integrated them with data held by MaizeGDB. This enables both exploring individual genomes, and comparing them in sets. In anticipation of more genomes expected in the near future, MaizeGDB developed a set of minimum standards for adopting a new genome assembly, designed templates for collecting essential metadata related to the genome and assembly, enforced naming conventions set out by the maize nomenclature committee, created documentation to help submit genome assemblies to GenBank, and developed a pipeline for loading new assemblies. All of this enables comparative analysis. In addition to bringing in new genome assemblies and providing the research community with means of improvement, MaizeGDB will continue stewardship of the B73 genome assembly and annotation, which is expected to remain the representative reference maize genome assembly for the foreseeable future. Multiple, high-quality genome assemblies and annotations integrated with trait, phenotype, and germplasm data, will improve researchers' ability to conduct trait and germplasm analyses and to choose appropriate germplasm for breeding programs.

P0841: Maize, Sorghum, Millet, Sugar Cane, and related
MaizeMine: A Data Mining Warehouse for MaizeGDB
Christine G. Elsik[1,2], Aditi Tayal[1], Deepak R. Unni[1], Hung N. Nguyen[1], Jack Gardiner[1], Justin Le Tourneau[1] and Carson M Andorf[3], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)Division of Plant Sciences, University of Missouri, Columbia, MO, (3)USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA
MaizeMine (http://maizemine.maizegdb.org), the new data mining warehouse for MaizeGDB, accelerates genomic analysis by enabling researchers without scripting skills to create and export customized annotation datasets merged with their own research data for use in downstream analyses. MaizeMine uses the InterMine data warehousing system to integrate genomic sequences and gene annotations from the B73_RefGen_v3 and B73_RefGen_v4 genome assemblies, Gene Ontology (GO) annotations, protein annotations (UniProt), protein families and domains (InterPro), homologs (Ensembl Compara) and pathways (CornCyc, KEGG, Plant Reactome). MaizeMine also provides database-cross references between genes of the AGPv3.21, AGPv4 and RefSeq gene sets, as well as pre-computed expression levels for all three gene sets based on RNAseq data from the *Zea mays* Gene Expression Atlas (NCBI BioProject PRJNA171684).
MaizeMine provides simple and sophisticated search tools, including a keyword search, built-in template queries with intuitive search menus, and a QueryBuilder tool for creating custom queries. The Genomic Region search tool executes queries based on lists of genome coordinates, and supports both B73_RefGen_v3 and B73_RefGen_v4. The List tool allows users to upload identifiers to create custom lists, perform set operations such as unions and intersections, and execute template queries with lists. When used with gene identifiers, the List tool automatically provides gene set enrichment for GO and pathways, with a choice of statistical parameters and background gene sets. MaizeMine is particularly useful for tracking gene identifiers across gene sets to facilitate meta-analysis. Query results can be downloaded in several formats (tab delimited, GFF3, Fasta, BED, JSON, and XML).

P0842: Maize, Sorghum, Millet, Sugar Cane, and related
RefSeq - A Curated Annotation of the Corn Genome
Brian Smith-White[1], Francoise Thibaud-Nissen[2] and Terence D. Murphy[2], (1)National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD, (2)National Center for Biotechnology Information (NCBI), Bethesda, MD
NCBI created Zea mays RefSeq transcript-protein pairs and the associated NCBI Gene record from the INSDC maize accessions. This is a publicly available resource consisting of a non-redundant collection of sequence records for RNA transcripts some encoding a protein; each with an associated NCBI Gene record. The transcript-protein pairs are curated by the NCBI staff. The curation is based upon public sequence records in the INSDC, RNA-seq records in SRA and the genome sequence submitted to INSDC. As a result there are differences between the annotation submitted with the genome sequence and the annotation generated by NCBI Eukaryotic Annotation Pipeline. These are 1) differences in the N-terminal and C-terminal extents of the encoded protein, 2) differences in the exon structure, and 3) presence of gene/transcript information independent of presence of sequence in the current genome.

P0843: Maize, Sorghum, Millet, Sugar Cane, and related
Building Maize High Quality, Chromosome-Scale, *de novo* Genome Assemblies by Scaffolding Next-Generation Sequencing Assemblies with Bionano Maps Generated with the New Direct Labeling Enzyme
Yang Zhang, Jian Wang, Henry B. Sadowski, Ernest Lam, Tom Wang, Warren Andrews, Thomas Anantharaman, Zongxiang Zhou, Carly Proskow, Amy Files, Michael G. Saghbini, Han Cao and Mark Borodkin, Bionano Genomics, San Diego, CA
Compared to human and a few other model organisms, genomic and genetic studies of plant species with complex genomes have lagged behind. Most of the economically important crops still lack a gold-standard reference genome assembly, crucial to understanding their biology. Plant genomes are often complex and highly repetitive, making generation of high-quality genome assemblies very costly if not impossible with next-generation sequencing (NGS) in the absence of long-range structural DNA information.
Bionano genome mapping, using nickase-based labeling, has been an indispensable tool for genome assembly in plants and animals. A new direct labeling enzyme and protocol has shown orders of magnitude improvement in contiguity, while also improving the amount of NGS data that can be scaffolded. This is achieved by the elimination of systematic double-stranded breaks that nickases introduce. The new labeling approach maintains the integrity of long DNA and allows the production of affordable, contiguous, and accurate chromosome-scale genome assemblies that can span most repeat regions.
Here, we present the workflow for direct labeling of genomic DNA of plants and animals for the Bionano Saphyr system and show some exemplary results on maize B73 genome assembly. With the new direct labeling enzyme and protocol, the *de novo* assembly produced very contiguous genome maps with an N50 of 99.5 Mbp, which covered the whole B73 reference across all 10 chromosomes. Scaffolding with a PacBio NGS dataset with a N50 of 1.18 Mbp generates a hybrid assembly with an N50 of >100 Mbp where >95% of the NGS sequences are anchored.

P0844: Maize, Sorghum, Millet, Sugar Cane, and related

Single Cell Transcriptomics: Perspectives from Maize

Jack Satterlee, Plant Biology, School of Integrative Plant Science, Cornell University, Ithaca, NY and Michael Scanlon, Cornell University, Ithaca, NY

Single cell RNA-Seq (scRNA-Seq) has emerged as a technology to facilitate the high-throughput transcriptomic analysis of individual cells. Numerous scRNA-Seq studies are reported in animals; however, the technology has seen limited use in plants. Here I present preliminary single cell transcriptomic data generated from maize embryos using the 10X Genomics Chromium™ platform. The maize embryo was selected as a study system owing to its diversity of cell and tissue types, including the embryonic shoot and root apical meristems (SAM and RAM, respectively). We aim to use this technology in combination with existing *in situ* hybridization and RNA-Seq data to examine the contributions of cell-type heterogeneity, signaling, and differentiation programs to SAM patterning and development, and to resolve single-cell gene co-expression networks.

Analysis of the dataset identifies embryo-specific marker genes, including previously-described lowly-expressed genes and transcripts accumulating in small cell populations. A preponderance of cells expressing epidermal cell marker genes were identified, suggesting biases in cell isolation. Dimensionality reduction resolves distinct cell clusters, but differential gene expression analysis indicates that cell variation in highly expressed housekeeping genes may explain some of these patterns. Nonetheless, marker gene expression analysis and pseudotemporal ordering of cells along differentiation trajectories reveals potentially pertinent developmental biology.

Future work will attempt to circumvent the unique challenges of unbiased cell-type isolation in plants by using nuclei as a source of RNA for transcriptomic profiling, and utilizing higher cell/nuclei populations to enhance statistical power.

P0845: Maize, Sorghum, Millet, Sugar Cane, and related

Comparing Alternative Splicing in Maize Using 4 Different Reference Genomes

Jerald D. Noble, University of Florida, Gainesville, FL

Maize is a highly polymorphic species with extensive genetic diversity, gene copy number variation, and gene presence-absence variation between genotypes of inbred lines. The current reference genome used for maize bioinformatics studies is the genome of the B73 inbred line. Previous work has been conducted investigating alternative splicing (AS) dynamics in maize between tissues, as a result of stress responses, and between genotypes.

AS is a process enabling multiple transcript isoforms to be coded from a single gene thereby expanding an organism's proteome without increasing the size of the genome. Given the genetic diversity of maize, this study profiles the diversity and conservation of AS between 27 maize genotypes. However, instead of comparing these AS events between genotypes using only the B73 v4 reference genome, this study integrates the AS events discovered in these genotypes relative to the W22, PH207, and CML247 reference genomes. Currently all AS events occurring in these genotypes relative to these reference genomes have been discovered and validated. Preliminary results generated by comparing the AS events between genotypes relative to each reference genome and by comparing the AS events discovered within genotypes using different reference genomes.

P0846: Maize, Sorghum, Millet, Sugar Cane, and related

Optimizing Genomic-Assisted Recurrent Selection for Small Breeding Programs in Developing Countries

Kebede T Muleta[1], Gael Pressoir[2], Edward S. Buckler[3] and Geoffrey P. Morris[1], (1)Department of Agronomy, Kansas State University, Manhattan, KS, (2)Chibas - Quisqueya University, Bon Repos, Croix des Bouquets, Haiti, (3)School of Integrative Plant Sciences, Section of Plant Breeding and Genetics, Cornell University, Ithaca, NY

The sorghum breeding program in Haiti, and other young breeding programs in developing countries, aim to improve the genetic potential of sorghum varieties under the challenge of limited resources, multipurpose yield targets and diverse agronomic environments. Implementing new genomic-assisted breeding method that exploits the recent developments in genomics and statistical methods to design novel breeding approaches could provide breeders the potential to

address the challenges of developing countries breeding programs. Here, we used simulations to test if genomic-assisted recurrent selection increases genetic gain compared to phenotypic recurrent selection in small new breeding programs. We investigate the levels of genetic gain, genetic diversity, and prediction accuracy resulting from genomic-assisted recurrent selection under various levels of population size per breeding cycle, selection intensity, trait genetic architecture, and GxE effects, in comparison with conventional recurrent selection. Our simulation results suggest that genomic-assisted recurrent selection will lead to a higher rate of genetic gain in small breeding programs by accelerating the breeding cycle, reducing the cost of population evaluation, and enabling evaluation of larger populations. Based on these positive findings, we have initiated genomic-assisted recurrent selection in the Haitian sorghum breeding program.

P0847: Maize, Sorghum, Millet, Sugar Cane, and related

Making Genomic Selection Possible in Low-Resources Breeding Programs

Gael Pressoir, Chibas - Quisqueya University, Bon Repos, Croix des Bouquets, Haiti, Jean Rigaud Charles, Chibas - Quisqueya University, Port au Prince, Haiti, Kebede T Muleta, Washington State University, Pullman, WA; Department of Agronomy, Kansas State University, Manhattan, KS, Ed Buckler, USDA-ARS, Cornell University, Ithaca, NY and Geoffrey Morris, Kansas State University, Manhattan, KS

Chibas (Quisqueya University) has had a very successful sorghum breeding program in recent years. We have released the *Papèpichon* variety that is resistant to the sugarcane aphid (*Melanaphis sacchari*). Farmers' adoption of this new variety has shown to be rapid and over half of the sorghum acreage in Haiti is already grown with *Papèpichon* less than a year after its release.

Chibas is now shifting its program to genomic selection. Here we report on the validation of the Chibas training set (all 252 lines developed to date) to make accurate genomic predictions and on the study of the population structure and genetic diversity of the lines developed to date compared to sorghum overall genetic diversity (see Poster #31318).

In less than a year our program as successfully switched to using genomic prediction and we have reduced the length of our selection cycle from three to one season (12 to 4 months). The empirical measurement of genetic gain per cycle and per year is under way between our traditional S1 recurrent selection scheme and the newly implemented genomic selection scheme.

P0848: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Dissection and G × E Analysis of Yield Component Traits Collected By Automated Image Analysis in a Collection of Diverse Maize Inbred Lines

Celeste Marie Falcon[1], Nicholas J Haase[2], Nathan Miller[1], Candice N. Hirsch[3], Marna D. Yandeau-Nelson[4], Edgar Spalding[1], Shawn Kaeppler[5] and Natalia de Leon[1], (1)University of Wisconsin-Madison, Madison, WI, (2)DuPont Pioneer, Mankato, MN, (3)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, (4)University of Iowa, Ames, IA, (5)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI

Studying yield component traits allows plant breeders and geneticists to further dissect the genetic architecture of yield and to exploit this knowledge to help meet the world's growing energy demands. We aimed to identify genomic regions that contribute to yield component traits in maize using genome-wide association analysis (GWAA) of 837 diverse maize inbred lines, which were genotyped with 430,948 RNA-seq based single nucleotide polymorphism (SNP) markers. We also sought to analyze genotype-by-environment interactions (G × E) for these traits. The lines were evaluated in replicated trials in 2013 and 2014 in WI. A subset consisting of 500 lines was evaluated in four environments in MN and one in IA in 2015 and 2016. An automated analysis pipeline was deployed for images of ears, cobs, and kernels to computationally extract nine phenotypic measurements from these images: ear width, ear length, kernel row number, kernels per row, kernel weight, kernel depth, kernel width, kernel area, and kernel thickness. G × E explained 5.60 to 11.35 percent of the variance in the trait values. Based on best linear unbiased predictor values calculated from all environments, between one and eight genomic associations were identified for each trait with some chromosome regions showing association with multiple traits. Conducting GWAA in each environment separately revealed few associations that were present for multiple locations, underscoring the influence of G × E on these traits. The significantly associated SNP markers, especially those identified for multiple environments, could be used in breeding approaches to develop higher yielding cultivars.

P0849: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Gains from Targeted Recombination in 27 Elite Biparental Maize Populations

Sofia P. Brandariz Zerboni, University of Minnesota. Department of Agronomy and Plant Genetics, Saint Paul, MN and Rex Bernardo, University of Minnesota. Department of Agronomy and Plant Genetics, Saint Paul, MN, Saint Paul, MN

Targeted recombination is the ability to induce or select for specific recombination points on the chromosomes of a species. A previous study with the intermated B73 × Mo17 maize (*Zea mays* L.) population suggested that targeted recombination may double the current gains for grain yield. Our objective in this study was to estimate genetic gain with targeted recombination compared to nontargeted recombination in elite maize germplasm. A total of 27 biparental maize populations were phenotyped at four to 12 environments in the U.S. from 2000 to 2008. The parents of the populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers, and marker data were imputed from lower-density screening of the progeny in each biparental cross. Targeted single- and double-recombination points were identified from genomewide marker effects. For the 27 populations, targeted recombination led to greater genetic gains than nontargeted recombination for each trait. Specifically, the relative efficiency (%) of selection with targeted single recombination versus selection with nontargeted recombination had a mean and range (in parentheses) of 253 (0, 485) for yield, 282 (0, 438) for moisture, and 281 (195, 469) for test weight. With targeted double recombination, the mean relative efficiencies increased to 326–368% across the three traits. We concluded that targeted recombination is a most promising breeding approach.

P0850: Maize, Sorghum, Millet, Sugar Cane, and related

Predicting Genetic Variance from Genomewide Marker Effects in Maize

Emmanuel Adeyemo, University of Minnesota, Saint Paul, MN and Rex Bernardo, University of Minnesota, St. Paul, MN

Predicting the genetic variance (V_G) in a biparental population has been difficult. An effective procedure for predicting V_G would likely require modeling of progeny segregation within each cross. Our objective was to determine whether the population mean, V_G and mean of the top 10% of progeny in a cross can be predicted effectively from genomewide marker effects. Eight maize (*Zea mays* L.) crosses that differed in their predicted mean and V_G were evaluated for plant and ear height, and growing degree days to silking across three locations in Minnesota in 2017. Each of the cross was represented by 120 to 144 random F_3 lines. Correlations between the observed and predicted means of each breeding population were significant (P = 0.05) for all three traits (0.93 for plant height, 0.77 for ear height and 0.80 for silking date). However, correlations between the observed and predicted V_G were non-significant ranging from -0.24 to 0.04 for the three traits. Correlations between the observed and predicted mean of the top 10% of progeny in each cross were significant for plant height (0.72) but not for ear height (0.55) and silking date (0.40). These results for predicting the mean of top 10% of progeny reflected the ability to predict the mean but not V_G. We concluded that while the means of breeding populations can be predicted effectively from genomewide marker effects, predicting the V_G of a cross remains a difficult task.

P0851: Maize, Sorghum, Millet, Sugar Cane, and related

Documenting the Manifestations of Heterosis in Co-Expression Networks for Maize

Peng Zhou[1], Peter Crisp[1], Sarah N. Anderson[1], Peter Hermanson[1], Steven Briggs[2] and Nathan M. Springer[1], (1)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, (2)University of California, San Diego, La Jolla, CA

Heterosis, or hybrid vigor, describes the phenomena in which hybrid offspring exhibit increased productivity or superiority compared to parental lines. A better understanding of how variation in parental lines combines to produce heterosis could be valuable for improving our ability to predict heterotic response or harness the potential of heterosis. While heterosis generally involves the interaction of multiple loci, many molecular analyses of heterosis have focused on single gene phenotypes. The comparative analysis of networks in hybrids relative to inbred parents may provide a more systematic way to evaluate the changes in regulatory patterns occurring during heterosis. Transcriptome

profiling was performed on 23 maize tissues across different developmental stages in three genotypes (B73, Mo17 and the B73xMo17 hybrid). Gene co-expression networks (GCN) were developed for each genotype. These networks were enriched for both Gene Ontology (GO) categories and CornCyc pathways suggesting the capture of biologically relevant information. An overall comparison of the three networks suggests that co-expression patterns are more conserved than the actual expression levels of genes. Modules (genes showing similar expression pattern) were identified in each network and the preservation of these modules was assessed in the hybrid compared to parents. In addition, co-expression relationships among genes annotated in the same pathways were evaluated in all three networks relative to permuted networks. Differentially expressed genes, genes with altered co-expression patterns as well as a number of hybrid-specific modules were identified. We hope to apply co-expression network analysis to develop a deeper understanding of the interactions and regulatory changes between the inbred parents and the heterotic hybrid.

P0852: Maize, Sorghum, Millet, Sugar Cane, and related
Assessing the Efficacy of Genomic Selection in the Improvement of Maize Maternal Haploid Inducers
Henrique Uliana Trentin[1], Ursula K. Frei[2] and Thomas Lubberstedt[1], (1)Iowa State University, Ames, IA, (2)Iowa State University/The DH Facility, Ames, IA
The production of doubled haploid lines in corn is based on the use of haploid inducers. Haploid inducers have the inherent ability to generate haploid kernels when combined with other maize plants. Even though haploid induction is caused by mutations on a pollen-specific phospholipase gene (MATRILINEAL - Kelliher, 2017), the rate of haploid induction is known to be affected by multiple genes (Lashermes and Beckert 1988, Deimling 1997, Röber 1999). Quantitative traits such as plant height, length and amount of pollen produced, as well as qualitative traits, such as stalk, root and kernel coloration are important traits for inducers. In this poster, we elaborate on different selection strategies that can be applied for inducer development. The efficacy of genomic selection on the improvement of haploid induction ability, plant height and length of pollen shed will be tested. A set of 196 inducer lines were genotyped and will be used as the validation population. Two sub-populations of 215 individuals with distinct degrees relatedness to the genotypes in the training population will be used as validation populations. Part of the training and the two validation populations will be phenotyped on the summer of 2018. Haploid induction ability will be evaluated by pollinating at least ten ears of a donor planted at three different times. The efficacy of genomic selection on the improvement of quantitative traits of haploid inducing lines will be evaluated.

P0853: Maize, Sorghum, Millet, Sugar Cane, and related
The History of Maize Domestication and Adaptation As Revealed By a Genome-Wide Survey of SNP Variation
Gen Xu, China Agriculture University, Beijing, China
Modern cultivated maize (*Zea mays* L. ssp. *mays*) has been heavily selected during domestication and adaptation. To better understand these processes, we conducted a genome-wide survey of 982 maize inbred lines and 190 teosinte accessions using over 40,000 SNP markers. Population structure, principal component analysis, and phylogenetic trees all consistently reflected historical evolutionary relationships among *Zea* species and subspecies. Shared haplotype analysis showed similar high levels of gene flow from *Z. mays* ssp. *parviglumis* and ssp. *mexicana*, confirming the critical contribution of ssp. mexicana to the maize gene pool. Scans for selection signatures identified 319 domestication sweeps and 406 adaption sweeps by analysis of wild, tropical and temperate maize, as well as with a set of some known genes such as *tb1*, *pbf1*. To verify the phenotypic effect of the selected regions, we compared the previous reported flowering time QTLs with selective sweeps, 196 domestication-selective sweeps and 238 adaption-selective sweeps were located within known flowering time QTL regions. Furthermore, a genome-wide association study on flowering time related traits were performed, 6 significant association signals were detected within the identified selective sweeps. Inverted chromosome segments provide an opportunity to look for evidence of natural selection, eight long-range inversions were detected based on unusual patterns of linkage disequilibrium (LD) in the wild *Z. mays* subspecies *parviglumis* and *mexicana*. All these results will provide insights into the evolutionary history of maize and be valuable for future maize breeding programs.

P0854: Maize, Sorghum, Millet, Sugar Cane, and related
Differential Contribution of Genomic Regions to Plant Genome Divergence
Jinyu Wang[1], Xianran Li[1], Jessica Barb[1], Michael Scanlon[2], Scott A. Jackson[3], Nathan M. Springer[4] and Jianming Yu[1], (1)Iowa State University, Ames, IA, (2)Cornell University, Ithaca, NY, (3)University of Georgia, Athens, GA, (4)Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN
The genome divergence pattern, modern accessions having significantly higher [AT] values than their wild progenitors, has been discovered across multiple species. However, the relative contribution of different genomic regions to, and the underlying mechanisms of this genome divergence pattern have not been well understood. Here, in maize and soybean, we observed higher [AT] values in domesticated accessions than wild accessions with genic and nongenic single nucleotide polymorphisms (SNPs), and nongenic SNPs have greater contributions to the A&T-increase. More intriguingly, the separation in [AT] is significantly enlarged in pericentromeric region for both genic and nongenic SNPs. The solar-UV-signature motifs (PyCG) are enriched around polymorphic sites in both species, which suggests UV radiation is likely one of the major forces driving the divergence of plant genomes. Surveying population-private SNPs further reveals the overrepresentation of solar UV-induced mutations in domesticated accessions. With base-composition across polymorphic sites as a genome phenotype, genome-wide scans identified a set of putative candidate genes involved in UV damage repair pathways. Research discoveries from this study will enrich our understanding of genome evolution.

P0855: Maize, Sorghum, Millet, Sugar Cane, and related
The Opaque-2 Regulatory Network in Maize Endosperm
Junpeng Zhan[1], Guosheng Li[1], Choong-Hwan Ryu[1], Shanshan Zhang[1], Chuang Ma[1], Alan Lloyd[2], Gary N. Drews[2], Xiangfeng Wang[1] and Ramin Yadegari[1], (1)University of Arizona, Tucson, AZ, (2)University of Utah, Salt Lake City, UT

Endosperm is a filial seed structure that provides nutrients and signals essential for embryogenesis and seedling germination. In contrast to dicotyledonous plants such as Arabidopsis in which the endosperm is eventually absorbed in part by the developing embryo, the endosperm in cereals persists through seed development and accumulates a high level of storage compounds, including starch and storage proteins. *Opaque-2* (*O2*) encodes a bZIP family transcription-factor protein, and has been shown to be a major regulator of seed storage-protein gene expression in maize. We used a coupled RNA sequencing (RNA-Seq) and chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) approach to identify genes directly or indirectly regulated by O2 in the maize endosperm. Using RNA-Seq, we identified 1,863 genes differentially expressed between the wild-type B73 inbred line and an *o2* mutant in B73 background. Among these, 186 genes were detected as putative direct O2 targets by ChIP-Seq. Analyses of the direct and indirect O2 targets revealed a broad role of O2 in the regulation of gene expression programs associated with multiple aspects of endosperm development and function. The identification of several direct target-associated *cis*-motifs that are presumably bound by O2 cofactors, and the observation that O2 directly activates two genes encoding O2 cofactors, provided novel insights into the key players in the O2 regulatory network. Furthermore, an analysis of the temporal expression patterns of O2 targets in WT vs. mutant endosperm revealed diverse modes of gene activation by O2.
This work was supported by National Science Foundation grants DBI-1261830 and IOS-1444568.

P0856: Maize, Sorghum, Millet, Sugar Cane, and related
Quantitative Trait Loci Analyses of Heterosis for Seedling Biomass-Related Traits in Maize Using Triple Testcross Population
Tifu Zhang[1,2], Shuaiqiang Liang[1], Feng Lin[1], Yuanda Lyu[1] and Han Zhao[1], (1)Jiangsu Academy of Agricultural Sciences, Nanjing, China, (2)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
Heterosis of biomass-related traits (BRTs) in maize has been emerged in seedling developmental stage. However, the underlying genetic basis has not been well understood yet. In this study, 122 IBMs (intermated B73 × Mo17) and 366 corresponding triple testcross (TTC) individuals obtained by crossing each IBM to its parents and F_1 were used to detect and characterize QTL for heterosis of four seedling BRTs: leaf length (LL), leaf width (LW), leaf area (LA), and seedling dry weight (SDW). The results showed that the SDW has the highest mid-parent heterosis (139.2%). Among seedling BRTs, the strongest significant correlation was found between LW and LA (R = 0.833). While, SDW has a significant but low R with other three seedling BRTs. QTL analyses were conducted using linear transformations Z_1, Z_2, and Z_3 calculated from means of TTC progenies. In total, 17 QTLs were detected for seedling BRTs in which 13 were augmented additive, two were augmented dominant, and two were dominance × additive epistatic. The contribution of a single QTL to total phenotypic variation ranged from 6.1% to 26%. Furthermore, four additive × additive digenic epistatic interactions were detected with the highest total R^2 for LW (41.0%), and two dominance × dominance digenic epistatic interactions were detected with the highest single R^2 for LA (21.2%). These findings display the complexity of the genetic basis for maize seedling BRTs and enhance our understanding on heterosis of maize seedling BRTs from the quantitative genetic perspective.

P0857: Maize, Sorghum, Millet, Sugar Cane, and related
Accurate Prediction of Grain Yield using its Contributing Genes for Enhanced Maize Improvement through Gene-Based Breeding
Meiping Zhang[1], Yanru Cui[2], Yun-Hua Liu[1], Wenwei Xu[1], Sing-Hoi Sze[3], Shizhong Xu[4] and **Hong-Bin Zhang**[1], (1)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (2)Department of Botany and Plant Sciences, University of California - Riverside, Riverside, CA, (3)Department of Computer Science and Engineering and Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, (4)Department of Botany & Plant Sciences, University of California, Riverside, CA
Many traits important to agriculture and human medicine are complex traits. Therefore, it is crucial to accurately predict the phenotypes of these traits for enhanced breeding in plants and animals and for enhanced medicine in humans. We previously cloned 1,501 *ZmINGY* (*Zea mays* inbred grain yield) genes significantly contributing to maize grain yield using a novel and genome-wide high-throughput gene cloning technology (*gExpress*). Here we report the accurate prediction of this complex trait using these *ZmINGY* genes, especially their number of favorable alleles (NFAs), genotypes and expression profiles. When their NFAs or genotypes were used for the prediction, only 27 of the genes gave a prediction accuracy comparable to those of genomic selection (GS) thus far achieved using thousands of DNA markers. When their expression profiles were used for the prediction, a prediction accuracy of $r = 0.85$ was achieved, approaching the maximal prediction accuracy. This prediction accuracy is higher than those of GS thus far achieved using genome-wide DNA markers by 63%. When two or all of the three *ZmINGY* datasets were jointly used for the prediction, the prediction accuracy of grain yield was 100% ensured. Therefore, we developed a gene-based breeding (GBB) system and a toolkit consisting of 150 key *ZmINGY* genes for enhanced grain yield breeding in maize and conducted GBB for grain yield using the three datasets of the toolkit. The results further demonstrated the ability, utility and efficiency of the *ZmINGY* for grain yield prediction, thus for enhanced and accelerated breeding in maize.

P0858: Maize, Sorghum, Millet, Sugar Cane, and related
Design Thinking and Data Mining in Plant Breeding
Jianming Yu, Tingting Guo and Xianran Li, Iowa State University, Ames, IA
Plant breeding is enhanced by integrating different scientific innovations and enabling tools. One major challenge that comes with the wide adoption of genomics and biotechnologies is to rethink and redesign the breeding programs at different stages and different scales. The essence of this new wave of breeding methodology research is to effectively identify and exploit genotype to phenotype relationship so that desirable cultivars are continuously and efficiently developed. Data mining, successful in many other areas, may provide solutions to address this question, particularly when findings are integrated into the design of new plant breeding pipelines. Enhanced by design thinking and data mining techniques, genomics-assisted prediction may reshape the plant breeding pipeline by enabling the efficient exploration of the enormous inference space of genetic combinations. We propose three essential components to streamline the breeding in the post-genomic era: better product creation (BPC), knowledge discovery from data (KDD), and optimal program design (OPD).

P0859: Maize, Sorghum, Millet, Sugar Cane, and related
Distinct Genetic Architectures for Phenotype Means and Plasticities in *Zea mays*
Aaron M Kusmec[1], Srikant Srinivasan[2], Dan Nettleton[1] and Patrick S. Schnable[1], (1)Iowa State University, Ames, IA, (2)IIT Mandi, Mandi, India
Phenotypic plasticity describes the phenotypic variation of a trait when a genotype is exposed to different environments. Understanding the genetic control of phenotypic plasticity in crops such as maize is of paramount importance for maintaining and increasing yields in a world experiencing climate change. Here, we report the results of genome-wide association analyses of multiple phenotypes and two measures of phenotypic plasticity in the maize nested association mapping (US-NAM) population grown in multiple environments and genotyped with ~2.5 million single nucleotide polymorphisms (SNPs). We show that across all traits the candidate genes for mean phenotype values and plasticity measures form structurally and functionally distinct groups. Such independent genetic control suggests that breeders will be able to select semi-independently for mean phenotype values and plasticity, thereby generating varieties with both high mean phenotype values and levels of plasticity that are appropriate for the target performance environments.

P0860: Maize, Sorghum, Millet, Sugar Cane, and related
Mapping QTLs for Kernel Row Number and *Fasciated* Ear By SNP-Based Bulk Segregant Analysis in Maize
Silvia Giuliani, Roberto Tuberosa and **Silvio Salvi**, DipSA - University of Bologna, Bologna, Italy
In maize, the number of kernel rows (KRN) of the ear is one of the most important grain yield components. Both QTLs and Mendelian mutations (such as abnormally shaped - *Fasciated* - ear mutants) have been discovered for this trait and utilized to gain information on the molecular genetic control of ear development. In this study, two Italian maize inbred lines were identified to show extreme phenotypes in terms of ear fasciation and low KRN, respectively and utilized to develop three recombinant inbred line (RIL) populations. Two of the populations (A and B) had the fasciated ear type inbred line as parent, while the third population (C) was generated by crossing the elite line B73 with the low KRN line. The three populations were thoroughly phenotyped for ear morphology and KRN in F5 and F6 generations and showed an overall continuous type of variation for ear traits. We next attempted to map QTLs for fasciated ear and KRN using bulk segregant analysis (BSA) based on a high-density maize SNP array (15k Illumina Infinium) in two successive years. Bulks included 15 plants (extremely fasciated ear plants or wild-type ear plants for populations A and B, and plants with highest or lowest KRN for population C). Preliminary results showed the presence of major QTLs segregating and affecting both ear fasciation and KRN.

P0861: Maize, Sorghum, Millet, Sugar Cane, and related
Detecting Candidate Genes Associated with Stalk Traits in Maize
Mona Mazaheri[1], Marlies Heckwolf[2], Brieanne Vaillancourt[3], Brett Burdo[2], Joseph Gage[1], Sven Heckwolf[2], Edgar Spalding[1], C. Robin Buell[3], Natalia de Leon[4] and Shawn Kaeppler[5], (1)University of Wisconsin-Madison, Madison, WI, (2)University of Wisconsin-Madison, (3)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, (4)UW, (5)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI
Corn stover is the largest source of crop residues in the United States and a promising sustainable energy source to replace fossil fuels. Stalk is the main component of stover, representing about half of the stover dry weight. To improve stover biomass as a biofuel feedstock, characterization of the genetic composition underlying stalk traits is essential. The objective of this study was to detect candidate genes associated with stalk anatomical traits (stalk diameter, rind thickness, vascular bundle area, and vascular bundle density), as well as plant height, using genome-wide association studies (GWAS). In doing so, a diverse panel of 942 maize inbred lines was phenotyped for stalk traits in replicated trials over multiple years. The population was genotyped with 899,784 RNA-Seq-based single nucleotide polymorphism (SNP) markers. GWAS were performed using a mixed linear model implemented within the FARM CPU package. Several candidate genes associated with stalk traits were detected. Some of the candidate genes, such as *FPA* and *Zmm22,* were flowering time genes and were associated with more than one stalk trait. Transgenic lines with high expression of *Zmm22* had a significant decrease in stalk diameter, plant height, below ear internode length, below ear node number, and ear size. In contrast, tassel length and above ear internode length were significantly increased in the *Zmm22* over-expression lines compared to the wild types. This study demonstrated the utility of the diversity panel for use in GWAS, revealed candidate genes associated with stalk traits, and uncovered common regulatory elements underlying multiple traits.

P0862: Maize, Sorghum, Millet, Sugar Cane, and related
Transcriptomic Analyses of Leaf Cuticular-Epidermal Development in Maize
Pengfei Qiao[1], Isabel Molina[2], Richard Bourgault[2], Laurie Smith[3], Michael A. Gore[4] and Michael Scanlon[1], (1)Cornell University, Ithaca, NY, (2)Algoma University, Sault Ste. Marie, ON, Canada, (3)UC San Diego, La Jolla, CA, (4)Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, Ithaca, NY
Cuticles serve important roles in protecting plants from water loss and pathogen attack. Composed of cutin and epicuticular waxes, cuticles are deposited on the epidermis of the plant shoot. To date, no transcriptomic study has explored the genes regulating cuticle biosynthesis along an expanding adult maize leaf. Biochemical profiling along the proximal-distal axis of the expanding eight leaf showed significant different wax and cutin contents, and the underlying transcriptomics are analyzed. We laser-microdissected the epidermis and non-epidermal tissues of leaf 8 in the inbred B73. After linear amplification of the tissue-specific RNAs, Illumina based RNA-seq identified transcripts implicated in cuticle development in maize. Differentially expressed genes were identified, and weighted co-expression network unravels the lipid biosynthesis genes and regulators correlated with the cuticular difference along the leaf. This study provides a transcriptomic perspective on cuticle development that is applicable to genetic strategies for drought tolerance in maize.

P0863: Maize, Sorghum, Millet, Sugar Cane, and related
Genetic Dissection of Leaf Angle Variation throughout the Maize Canopy
Matthew J. Dzievit, Xianran Li and Jianming Yu, Iowa State University, Ames, IA

Improved canopy architecture is one of the ways maize hybrids have adapted to higher plant densities. Modern hybrids have been moving increasingly towards upright leaves, which assist in distributing light more effectively in the canopy. This study is being conducted to identify genetic factors controlling leaf angle (LA) variation throughout the canopy. PHW30 has upright LAs throughout the canopy (74°-79°), while Mo17 has flatter LAs (62°-55°). B73 has upright LAs in the upper part of the canopy (79°), and flatter LAs in the lower canopy (62°). Four reciprocal bi-parental populations with PHW30 as the common parent were developed, and LA in the lower canopy of the F_2 and $F_{2:3}$ progeny was measured. Genotyping by sequencing was used to genotype, and quantitative trait loci (QTL) mapping was conducted for each trait and population combination. Three QTLs were identified in both the F_2 and $F_{2:3}$ generations explaining 33% of the phenotypic variance for the B73 populations. Two QTLs were identified in both the F_2 and $F_{2:3}$ generations explaining 40% and 30% of the phenotypic variance for the Mo17 populations. QTLs on chromosome one and three were identified across populations and generations. Additionally, selected progeny from the F_2 populations were used to generate double haploid lines, and measure LA at four leaf positions throughout the canopy. QTL mapping will be done using the four LA positions. Identification of QTL controlling LA for specific leaves will be useful for optimizing LA throughout the canopy to maximize light interception and further increase yield.

P0864: Maize, Sorghum, Millet, Sugar Cane, and related

Engineering C4 Photosynthesis in Maize to Enhance Nitrogen Utilization

Jennifer Arp[1], Doug Allen[1], Stephen P. Moose[2] and Thomas P. Brutnell[3], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)University of Illinois, Urbana-Champaign, Urbana, IL, (3)Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO

Crop breeding efforts will need to increase yields while decreasing the use of non-renewable resources and minimizing detrimental impacts on the soil, water, and air in order to continue to provide food for an increasing world population. Modifications to photosynthesis present an opportunity to increase crop yields. Maize and other C_4 plants evolved a more efficient type of photosynthesis where initial carbon fixation and Calvin cycle activity are spatially separated into the mesophyll and bundle sheath, in order to reduce losses associated with photorespiration. Three subtypes of C_4 photosynthesis exist, and maize utilizes the NADP-malic enzyme (NADP-ME) pathway in combination with the phospho*enol*pyruvate carboxykinase (PEPCK) pathway. In the PEPCK pathway, aspartate is used as a transfer molecule between mesophyll and bundle sheath cells, and also acts in nitrogen metabolism, linking the two pathways. *Ds* insertion mutations were identified in the maize PEPCK1 gene, and the pepck1-*Ds* plants were grown under low and high nitrogen in a nitrogen responsive field site during the summers of 2016 and 2017 to determine the effect of the PEPCK pathway under nitrogen stress. At the V8 growth stage, tissue was sampled along the developmental gradient of the leaf, and RNAseq was performed to determine the extent of compensation from the NADP-ME pathway in the mutant. Among agronomic traits, pepck1-*Ds* mutants flowered later, were taller, and had smaller kernels and heavier cobs than controls. The vegetative tissues contained more nitrogen and retained more biomass, indicating that pepck1-*Ds* plants were deficient in nitrogen and sugar remobilization.

P0865: Maize, Sorghum, Millet, Sugar Cane, and related

Discovery of Maize "Nitrogenes" – Regulatory Variants That Improve Nitrogen Utilization in Maize Hybrids

Stephen P. Moose[1], **Jessica M Bubert**[1], Jennifer Arp[2], Yuhe Liu[1], Brian H Rhodes[1], Christine J. Lucas[1], Bosen Zhang[1] and Devin Nichols[1], (1)University of Illinois, Urbana-Champaign, Urbana, IL, (2)Donald Danforth Plant Science Center, St. Louis, MO

Our research group has investigated for the past decade the genetic basis for the tremendous growth response of maize hybrids to nitrogen (N) fertilizer. This research program has developed the infrastructure to monitor N use phenotypes at field and population scales, leveraged community resources for functional genomics research, and exploited novel genetic variation for N use traits. Important conclusions include: 1) the gene regulatory programs that respond to nitrogen in inbreds are distinct from those operating in hybrids, 2) variation for N utilization phenotypes in maize hybrids is associated with combining diversity in regulatory programs, and 3) N-responsive regulatory programs include both genetic and epigenetic components. This information has enabled the discovery and validation in field trials of multiple genetic variants that modulate nitrogen utilization in maize hybrids. These variants influence key traits of N acquisition by roots, root architecture, N remobilization from vegetative source to reproductive sink tissues, harvest index, and N storage versus growth during seed filling. For many "NitroGenes", heterozygosity is critical to increasing N utilization, which has important implications for gene discovery and improvement of N use efficiency in other crops.

P0866: Maize, Sorghum, Millet, Sugar Cane, and related

Transcriptional Response of Maize Carbonic Anhydrase Mutants to CO2 Limitation

Allison R. Kolbe[1], Anthony J. Studer[2], Omar Cornejo[1] and Asaph B. Cousins[1], (1)Washington State University, Pullman, WA, (2)University of Illinois, Urbana, IL

Carbonic anhydrase (CA) catalyzes the hydration of CO_2 in the first biochemical step of C_4 photosynthesis, and has been considered a potentially rate-limiting step in monocots. Previous work generated a double knockout of CA1 and CA2 in maize, reducing total leaf CA activity to less than 3% of wild-type. Analysis of these mutants demonstrated that CA does not limit photosynthesis in maize at ambient or higher CO_2 concentrations. However, these CA mutants exhibited reduced rates of photosynthesis at sub-ambient CO_2, and accumulated less biomass when grown at 100 ppm CO_2. In order to clarify the importance of CA for supplying HCO_3 for C_4 photosynthesis, RNA-seq was performed on *ca1* and *ca1ca2* mutants. Mutants and wild-type plants were grown at 10,000 ppm CO_2 and transitioned to 100 ppm CO_2. Samples were taken at high CO_2 and at two time points after the low CO_2 transition, in order to identify immediate and longer-term responses to CO_2 deprivation. Despite the existence of multiple isoforms of CA, no other CA genes were upregulated in CA mutants. Although photosynthetic genes were downregulated in response to low CO_2, differential expression was not observed between genotypes. However, multiple indicators of carbon starvation were present in the mutants, including amino acid synthesis, carbohydrate metabolism, and sugar signaling. Furthermore, these data support a role for CA in stomatal signaling. This experiment provides insight into the biological function of CA as well as the gene expression response of maize to changes in CO_2 concentrations.

P0867: Maize, Sorghum, Millet, Sugar Cane, and related

Allelic Imbalance Expression of Long Noncoding RNAs in Maize Hybrid Under Nitrogen Stress

Yuanda Lyu, Shuaiqiang Liang, Min Ge, Ling Zhou, Bing He, Tifu Zhang and Han Zhao, Jiangsu Academy of Agricultural Sciences, Nanjing, China

Although previous studies have evaluated the allele-specific expression(ASE) analysis of coding genes and those impact on heterosis, the imbalanced regulated effect of long non-coding RNAs in maize hybrid has been unknown. Here we employed allele-specific exression (ASE) analysis against long non-coding RNAs to evaluate their imbalanced allelic expression under Nitrogen stress in maize hybrid. By integration with PacBio fl-cDNA sequencing and ssRNA-Seq technology, a set of full-length transcripts including coding RNAs and non-coding RNA were generated and their expression profiles were tested and normalized by Tuxedo pipeline. Nitrogen-response lncRNAs were then identified and ASE analysis was performed in hybrid lines under Nitrogen stress. Cis- and trans-regulatory divergence were found at informative single-nucleotide polymorphism sites of long non-coding RNAs, respectively. some sites exhibiting both cis- and trans-effects. At last, Co-expression network analysis between imbalanced expressed coding genes and non-coding genes were performed for predicting the potential role of those N-response imbalanced expressed lncRNAs. The study will provide a unique insight for heterosis.

P0868: Maize, Sorghum, Millet, Sugar Cane, and related

Cambridge India Network for Translational Research in Nitrogen (CINTRIN)

Rajeev Gupta[1], Rakesh K. Srivastava[1], Santosh Deshpande[1], Abhishek Rathore[1], Girish Chander[1], Kirandeep Romana[1], Srikanth Bollam[1], P Vijayalakshmi[1], Manoj Prasad[2], Parveen Chhuneja[3], Satinder Kaur[3], Varinder P Singh[3], Ottoline Leyser[4], Howard Griffiths[5], Alison R Bentley[6], Emma Wallington[7] and Tina Barsby[7], (1)ICRISAT, Patancheru, India, (2)National Institute of Plant Genome Research, NEW DELHI, India, (3)Punjab Agricultural University, Ludhiana, India, (4)Sainsbury Laboratory at Cambridge University (SLCU), Cambridge, United Kingdom, (5)Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom, (6)NIAB, Cambridge, United Kingdom, (7)National Institute of Agricultural Botany (NIAB), Cambridge, United Kingdom

The global demand for nitrogen (N) fertilizer for agricultural production, which already stands at ~110 million metric tons per year, is projected to increase to ~250 million metric tons by the year 2050. A substantial amount of N applied for crop production is lost by leaching, run-off and de-nitrification which not only pollute the ground water and adversely affect soil structure but also has detrimental effects on environment such as increase in nitric oxide, ozone etc. Hence, developing crop varieties with improved N use efficiency will help mitigate these problems to some extent. The CINTRIN, a consortium of 8 partners (NIAB, UCAM, SLCU, ADAS, KisanHub from UK and ICRISAT, PAU, NIPGR from India), is funded by BBSRC/Newton Fund in UK and DBT in India. The overarching aim of CINTRIN is to improve not only the income and livelihood of farmers by reducing the inputs cost, but also to save the environment by minimizing the negative impacts of excessive use of fertilizers. The natural variation for N use efficiency will be studied in diverse germplasm of wheat, sorghum, pearl millet and foxtail millet. The findings will be applied to develop new breeding lines with enhanced NUE. CINTRIN will also use model plants such as *Arabidopsis* and *Brachypodium* for basic research which will be translated into crops in future. In addition, CINTRIN will prime the long term scientific relationship of ICRISAT and other Indian partners with those in the UK by exchange visits of scientists/students, scientific meetings, workshops and exchange of technologies.

P0869: Maize, Sorghum, Millet, Sugar Cane, and related

In-Situ QTL Mapping of Foliar and Tassel Heat Stress Resistance in Maize

James P McNellie[1], Junping Chen[2] and Jianming Yu[1], (1)Iowa State University, Ames, IA, (2)USDA-ARS, Lubbock, TX

Genetic studies into heat stress resistance using field grown *Zea Mays* are few. This is problematic as climate change will increase the frequency and severity of heat stress events. To meet the challenges of feeding a growing population, a better understanding of the genetic architecture and molecular mechanisms controlling heat stress resistance in maize is required. Foliar and tassel heat stress phenotypes were scored in two recombinant inbred line populations, B73 x NC350 and B73 x CML103, for two to three years (2010 – 2012) in Lubbock, TX. Irrigation ensured well-watered conditions to remove confounding effects of drought. When possible, foliar traits were scored at three vegetative developmental stages; early (before V10), middle (V10 – V14) and late (tassel emergence, VT). Phenotyping occurred following a heat stress event, defined as three consecutive days with maximum air temperature greater than 36°C. There were 22 significant QTL detected with little difference in QTL number and position across developmental stages. With few exceptions, QTL for different stress phenotypes mapped to unique regions. This study represents preliminary steps to better understand the genetic control of heat stress resistance in maize.

P0870: Maize, Sorghum, Millet, Sugar Cane, and related

Toward Cloning a New Maize *tasselseed* Mutant

Silvio Salvi, Riccardo Bovina, Silvia Giuliani, Elisabetta Frascaroli, Roberto Tuberosa and Pierangelo Landi, DipSA - University of Bologna, Bologna, Italy

Unisexual maize flowers originate through selective abortion of female primordia in the tassel and of male stamens in the ear from bisexual inflorescences. *Tassel seed* mutations are known to alter the usual sex fate allowing carpel survival in the male inflorescence. Objective of the present research is to describe and map a novel *tassel seed* phenotype shown by an inbred line, Rig7, identified among a set of lines derived from *in vitro* regeneration. Genetic mapping was carried out using a B73 x Rig7 F_2 population (genotyped with 15K SNP array) and by SNP-based bulk-segregant-analysis using two additional populations (BC_1 and F_2). Both approaches clearly indicated that the *tassel seed* phenotype is under the control of two loci mapping on chromosomes 2 and 6. A strong and unexplained reduction in recombination across chromosome 2 precluded the characterization of the locus on such chromosome. On the contrary, the locus on chromosome 6 was mapped to a < 2 Mb region on bin 6.07. Further fine mapping analysis using 2,000 F_2 recombinants and corresponding F_3 and F_4 families enabled us to narrow the *tassel seed*-6.07 locus to a 130-kb region which included three genes based on B73 genome annotation. The role of different candidate genes are being tested by comparison of allele sequences and by testing the concentration of different hormones in young tassel sample tissues.

P0871: Maize, Sorghum, Millet, Sugar Cane, and related
Analysis of Yellow Striped Mutants in *Zea mays*
David Chan-Rodriguez, University of Massachusetts Amherst, Amherst, MA
Iron-deficiency anemia is one of the most prevalent forms of malnutrition worldwide affecting more than 1.5 billion people, with children and pregnant women being the most affected groups. Iron (Fe) is equally important in plants to perform essential functions such as photosynthesis. Grasses, which include most of the world's staples, acquire iron from the soil by secreting chelator molecules called phytosiderophores (PS), to bind iron. Roots then transport PS-Fe into root cells using the transporter YS1. By elucidating key elements in iron homeostasis in grasses, new strategies for biofortication can be elaborated. In this study, we aim to bring to light new genes involved in iron homeostasis by characterizing 32 "yellow striped" mutants from the Maize Genetics Cooperation Stock Center. By performing complementation test between the maize mutants *ys1*, *ys3* and these unknown mutants, new genes affecting iron utilization may be identified. Moreover, by identifying and analyzing new *ys3* alleles, we aim to determine whether the suggested candidate gene ZmTOM1 underlies the long-known *ys3* mutant. We found 3 novel yellow striped mutants, that are not allelic to *ys1* or *ys3*. In all these *ys** mutants, Fe content in leaves is significantly lower than WT indicating that the yellow striped phenotype is a symptom of iron deficiency. We are evaluating whether *ys** mutants represent one or multiple genes by performing complementation test among them. In addition, we identified 4 new *ys3* alleles in which *ZmTOM1* exonic region was analyzed to identify causative mutations. We found evidence that *ys3* gene is *ZmTOM1*.

P0872: Maize, Sorghum, Millet, Sugar Cane, and related
Mapping Pericarp Anthocyanin Content in a Maize Landrace Apache Red
Laura A. Chatham and John A. Juvik, University of Illinois, Urbana, IL
Interest in natural alternatives to artificial food colorants has prompted the search for an economical source of natural pigment. Maize represents a scalable and efficient source of anthocyanins, natural orange to blue pigments, capable of meeting growing market demands. Lines containing pigment in pericarp, the outermost layer of the kernel, often yield anthocyanin concentrations an order of magnitude greater than the more frequently studied maize lines producing anthocyanin in aleurone tissue, the outermost layer of the endosperm. Here we investigate diversity in anthocyanin content and the corresponding genetics in 176 lines derived from Apache Red, a maize landrace with pigmented pericarp. Extracts varied widely in concentration of cyanidin-, pelargonidin-, and peonidin-based anthocyanins. Principal component analysis and hierarchical cluster analysis revealed several unique anthocyanin compositional clusters. Many lines contained flavanol-anthocyanin condensed forms, pigments not frequently found in aleurone pigmented lines, but that influence pigment hue and stability. We genotyped each line using Illumina sequencing, and performed association mapping to identify loci responsible for the observed variability in anthocyanin content. To maximize signal detection, we used a model that excludes markers on the same chromosome as the tested marker when calculating relatedness, allowing the discovery of promising candidate genes associated with pericarp anthocyanin content. The lines developed and analyzed herein will serve as a valuable source of diversity both for understanding the genetics underlying anthocyanin biosynthesis in maize pericarp and for breeding lines capable of replacing synthetic dyes.

P0873: Maize, Sorghum, Millet, Sugar Cane, and related
Domestication and the Response to Competition: Similarities and Differences between Corn and Teosinte
David P. Horvath, USDA-ARS, Fargo, ND, Stephanie Bruggeman, South Dakota State University, Brookings, SD and Sharon Clay, South Dakota State University, Brooking SD, SD
Crops respond to weed pressure by altering their transcriptomes. Corn has undergone significant selective pressures during domestication- first for ability to thrive under intercropping conditions for growth with beans and squash, and later for high density planting under intense intra-species competition. We hypothesize that these selective pressures have altered the response of corn to competition. We have previously examined the transcriptomic response of corn to weed competition using RNAseq technologies, and here we present similar studies on teosinte with a comparison to what we observed with corn. Two different varieties of teosinte were grown with or without weeds and harvested at a time that co-planted nearby corn had reached the V8 stage of development in four replicate plots. Previously, corn was also grown with or without weeds to a similar developmental stage in two consecutive years (also in four replicate plots each year). RNAseq analysis was done on three of the four plots for each experimental run using samples of RNA collected from the distal 4-6 inches of the upper-most leaf. The resulting sequences were assembled *de novo* with Trinity and subjected to differential expression analysis using the RSEM program suite. To avoid detecting any species-specific allelopathic responses, the weed species were different in each year of the corn analysis and only those responses that were common between years were considered. In both corn and teosinte, gene set and subnetwork enrichment analyses indicated an increase in expression in genes associated with biotic stress responses involving salicylic acid, jasmonic acid, and auxin signaling and phosphate starvation. Also, some indications of dehydration stress were observed in both species. Ontologies associated with nitrogen utilization, growth and development, and photosynthesis were down regulated in both. Interestingly, ontologies associated with up-regulation of phytochrome signaling were over-represented in corn growing with weeds, but not with teosinte. Cytokinin-associated ontologies seem to be slightly more prevalent among genes down-regulated in teosinte weed responses relative to corn.

P0874: Maize, Sorghum, Millet, Sugar Cane, and related
Structural Basis of the Antifungal Activity of SUGARWINs Proteins and Their Role in Plant Defense
Flávia P. Franco[1], Renata O. Dias[1], Danyelle Toyama[2], Adelita C. Santiago[2], Flávio Henrique-Silva[2], Daniel S Moura[1] and Marcio C Silva-Filho[1], (1)Universidade de São Paulo - USP, Piracicaba, Brazil, (2)Universidade Federal de São Carlos - UFSCar, São Carlos, Brazil
Plants respond to insect attack by inducing and accumulating a large set of defense proteins. We identified two homologues of a barley wound-inducible protein (BARWIN) in sugarcane, which were designated SUGARWIN1 and 2 (sugarcane wound-inducible proteins). Although BARWIN function has not been fully established, antifungal properties have been described for a number of homologues. SUGARWIN1 and 2 genes expression are induced in response to wound and *Diatraea saccharalis* damage. Although the recombinant SUGARWIN protein does not affect insect development, it promotes significant morphological and physiological changes in *Fusarium verticillioides* and *Colletotrichum*

falcatum, which lead to fungal cell death via apoptosis. In this study, we deepen our understanding of the role of SUGARWINs in plant defense and the molecular mechanisms by which these proteins affect fungi by elucidating their molecular targets. We demonstrated that SUGARWINs are also induced by *C. falcatum* infection in sugarcane, and the induction of SUGARWINs can vary among sugarcane varieties. The sugarcane variety exhibiting the highest level of SUGARWIN induction exhibited a considerable reduction in *C. falcatum* infection. Furthermore, SUGARWIN1 exhibited ribonuclease and chitinase activity, whereas SUGARWIN2 exhibited only chitinase activity. This variable enzymatic specificity seems to be the result of divergent amino acid composition within the substrate-binding site that was demonstrated by protein modelling and docking studies. We confirmed this result producing mutants with altered active site and performing comparative analysis with the native protein. Our results show that SUGARWINs play an important role in plant defense against opportunistic pathogens and can be important to red rot disease control.
Financial Support: This work was supported by a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and by a Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

P0875: Maize, Sorghum, Millet, Sugar Cane, and related
Development of a Marker Assisted Sorghum Breeding Program
Terry Joseph Felderhoff, Kansas State University, Manhattan, KS and Geoffrey Morris, Department of Agronomy, Kansas State University, Manhattan, KS
Molecular markers are a powerful tool which can reveal genetic information and genomic structure which is unobtainable through traditional phenotypic observation. However, marker assisted selection has been historically underutilized as a breeding tool in the public sector of sorghum breeding. Here we present our work to implement an integrated marker assisted breeding program for sorghum. The first project we have undertaken for this new program is the integration of chilling tolerance from exotic germplasm into elite material. Chilling tolerance in sorghum was identified in Chinese germplasm almost 60 years ago, but high linkage disequilibrium between chilling tolerance loci and undesirable genes for US grain sorghum (grain tannins, tall stature, etc.) halted integration of these genetic traits. For this project, QTLs identified in a US × Chinese mapping population were used for the start of marker assisted crossing, intercrossing, and backcrossing. The data from the mapping population confirmed that the two largest-effect chilling tolerance QTL on chromosomes 4 and 9 were within 1 cM of the *Tannin1* and *Dwarf1* genes, respectively. With the aid of Kompetitive Allele-Specific Primer (KASP) PCR markers developed for the chilling tolerance QTL and the cloned *Tannin1* and *Dwarf1* genes, individual lines have been identified that contain rare recombinations between the chilling tolerance QTL and the undesirable genes which would have been almost undetectable through pure phenotyping. Integration of the QTL is underway, as well as production of near-isogenic lines (NILs) for fine-mapping the chilling tolerance QTLs.

P0876: Maize, Sorghum, Millet, Sugar Cane, and related
Initial Characterization of Selected Sorghum Phenotypes from a Publically Available EMS-Induced Mutant Population
Nicholas Gladman[1], Yinping Jiao[2], Gloria Burow[3], John Burke[3], Veronica Acosta-Martinez[3], Doreen Ware[4] and Zhanguo Xin[5], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)USDA-ARS/Cold Spring Harbor Laboratory, Lubbock, TX, (3)USDA-ARS, Lubbock, TX, (4)USDA/ARS - Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (5)USDA ARS, Lubbock, TX
Nick Gladman, Cold Spring Harbor Laboratory
Sorghum is an important global food, forage, and bioenergy crop with an annual production value of over two billion dollars in the USA. This C4 photosynthetic grass has resilience to heat, drought, low soil fertility, and other abiotic and biotic stresses that have made it an agricultural mainstay across geographic and socioeconomic boundaries. Through an EMS mutagenesis screen performed and curated by the USDA-ARS, hundreds of mutations were identified and determined to be causative for a diverse array of agriculturally useful phenotypes. Specifically, independent mutations were identified in several Gibberellic Acid (GA) biosynthesis and signaling genes that result in dwarf plants. Separately, other mutations were shown to be causative for decreased epicuticular wax (bloom) content on the culm and leaves. These observed mutations include missense, nonsense, and splice-site changes and correspond to unique expression patterns of genes within their respective molecular pathways. All of these mutations and their corresponding phenotypes represent strong candidates for integration into sorghum breeding programs for the purpose of improving grain yield and bioenergy and enhancing global agricultural security.

P0877: Maize, Sorghum, Millet, Sugar Cane, and related
Diversity Analysis of Sorghum Genotypes using SSR Markers, Agro-Morphological and Nutritional Quality Traits
Maletsema Alina Mofokeng, Agricutural Research Council-Grain Crops, Potchefstroom, South Africa and Kingstone Mashingaidze, Agricultural Research Council-Grain Crops, Potchefstroom, South Africa
Sorghum is one of the most important cereal crops grown for human and animal consumption. Knowledge of genetic diversity among the genotypes grown is essential for effective breeding. The objective of this study was to determine the genetic diversity present among selected South African sorghum genotypes using SSR, agro-morphology and nutritional quality traits. Hundred and three genotypes were analysed using 30 SSR markers, agro-morphological traits, protein and amino acid composition. The size and number of alleles ranged from 90 to 294 bp, and 2 to 15, respectively. The polymorphic information content varied between 0.02 and 0.84. The heterozygosity values ranged between 0.02 and 0.85, with the genetic distances ranging between 0 and 8.4. The principal component analysis (PCA) revealed three most important components contributing 38.9%, 30.96% and 18.13% to the total variation. Nineteen genotypes with high crude protein content were selected and analysed for amino acid profiles using protein hydrolysates. The crude protein content varied from 7.69 to 16.18% across the two sites with a mean of 13.07%. The genotypes that had high crude protein content at both sites were Mammopane, AS16 M1, Macia-SA, AS19, Maseka-a-swere, and AS4. The genotype AS16cyc was the best candidate for high phenylalanine content at 5.99%. High lysine was detected in genotypes 52461.1.1.1 at 2.27%, AS17 at 2.25%, Manthate at 2.16% and 1481.1.1.1 at 2.11%. Overall, the studied lines had great variability at DNA level, morphologically and in their protein and amino acid profiles, which will be useful for breeding.

P0878: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Dissection of Sorghum Dynamic Traits with Unmanned Aircraft System Based High-Throughput Phenotyping

Zhenbin Hu[1], Xu Wang[2], Jesse Poland[2] and Geoffrey Morris[1], (1)Department of Agronomy, Kansas State University, Manhattan, KS, (2)Department of Plant Pathology, Kansas State University, Manhattan, KS

Genetic architecture of dynamic traits of plants in the field is poorly understood due to the limited resolution of traditional field phenotyping. High-throughput phenotyping with unmanned aircraft systems (UAS) should facilitate genetic dissection of dynamic quantitative traits loci in crop field experiments. In the present study, UAS and low-cost multispectral camera was employed to collect imagery of a sorghum biomass association panel (BAP) 2–3 per week throughout the field season. A pipeline was developed for image-processing to extract plant phenotypes at plot level from the UAS images, and UAS-based phenotypic data with high heritability and strong correlations with manual data. The UAS data was used to develop a three-parameter logistic model for sorghum growth. Genomic regions underlying the phenotypic variation were identified over the growing season using genome-wide association study by combining the growth model as phenotypic data and whole-genome resequencing data. We found the dynamic QTL precisely colocalized with major height genes *dw1*, *dw2*, and *dw3* appeared at specific growth stage. Novel QTL on chromosome 1 and 9 were identified at early growth stage. This study shows that the UAS-based high-throughput phenotyping platform is reliable, and would facilitate the genetic dissection of dynamic traits in sorghum.

P0879: Maize, Sorghum, Millet, Sugar Cane, and related

Morphometric Diversity and Comparative Development of Sorghum Inflorescences

Samuel Leiboff, UC Berkeley / Plant Gene Expression Center, Albany, CA and Sarah Hake, UC Berkeley, Albany, CA

Sorghum bicolor (L. Moench) is a drought-resistant relative of maize that generates a highly-branched inflorescence of perfect flowers called a panicle. Although a great deal is known about the genetic pathways that regulate the formation of the maize tassel, less is known about the sorghum panicle. Using a high-throughput imaging platform I analyzed the panicles from a 200-inbred subset of the sorghum association mapping panel (SAP). Using publicly-available SNPs, I used a mixed-model GWAS to identify panicle morphology candidate genes. Natural diversity in sorghum panicle morphology did not identify orthologs of known, classical master regulators of maize tassel development. Instead, panicle morphology candidates implicated unexpected hormone transporters and transcription factor families. Further comparison of ontological events in sorghum panicle and maize tassel development by scanning electron microscopy (SEM) and synchrotron radiation microCT suggest that there may be many developmental differences in panicle and tassel formation. Ongoing transcriptomic and reverse genetic studies continue to explore the similarities and differences of inflorescence development in these closely-related species.

P0880: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic Prediction and Genetic Diversity in a Sorghum Population Undergoing Simultaneous Selection for Grain and Stem Sugar Yields

Jean Rigaud Charles[1], Kebede T Muleta[2], Geoffrey Morris[2] and **Gael Pressoir**[3], (1)Chibas - Quisqueya University, Port au Prince, Haiti, (2)Department of Agronomy, Kansas State University, Manhattan, KS, (3)Chibas - Quisqueya University, Bon Repos, Croix des Bouquets, Haiti

Chibas recently (2013) established its breeding program on sorghum; we have been breeding simultaneously for grain and stem sugar yields (dual purpose sorghum). Two cycles of recurrent selection (on S1 families) have been completed, and a total of 252 lines have been developed, to date. Chibas is now shifting its program to using genomic predictions in order to accelerate the rate of genetic gain per cycle and per year. Here we report on (1) initial results from the evaluation of the training set (all lines developed to date) and the accuracy of genomic predictions made in this population; and (2) on the genetic diversity and population structure of the Chibas population compared to a large sorghum diversity panel.

P0882: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic Signatures of Seed Weight Climate Adaptation in the Globally Genomic Diversity Panel of Sorghum Bicolor

Jianan Wang[1], Narinder Singh[2], Zhenbin Hu[1], Hari D Upadhyaya[3] and Geoffrey Morris[1], (1)Department of Agronomy, Kansas State University, Manhattan, KS, (2)Kansas State University, Manhattan, KS, (3)ICRISAT, Hyderabad, India

Sorghum is an important cereal crop in the world's drier regions with multiple botanical races morphological adaptations to diverse agroclimatic zones. Seed weight, a key component of grain yield, is a quantitative trait that varies among sorghum landraces from different agroclimatic zones. However, the genetic basis of seed weight variation in sorghum remains largely unclear, particularly in the relationship between seed weight and local adaptation or natural selection. The association between seed weight and botanical races suggest that genetic architecture underlying sorghum seed weight was shaped by agroclimatic adaptation. We investigated the genomic diversity and population structure of ~2000 georeferenced sorghum accessions collected globally using ~404,000 single nucleotide polymorphisms (SNPs) markers. We characterized correlations between 100-seed weight and environment variables. We also performed genome-wide association studies (GWAS) to identify the loci controlling the 100-seed weight, and identified the genomic signatures involved in local adaptation. This study indicated that environmental factors explain a substantial portion of seed weight variation, according to the statistical analyses. Moreover, population structure is highly related to the distribution of globally climatic zones. In addition, GWAS localized several candidate genes for 100-seed weight. Finally, a set of outliers loci (including yield-related loci) were identified, which were considered as the adaptive and selection signatures on the sorghum genome. Our new findings deepened the understanding of the genetic mechanism of seed weight and the effect of local adaptation to seed weight variations of sorghum.

P0883: Maize, Sorghum, Millet, Sugar Cane, and related

Identification and Characterisation of Candidate Genes for Fertility Restoration in Sorghum

Joanna Melonek, University of Western Australia, Crawley, Australia, David R Jordan, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Warwick, Australia, Emma S Mace, Centre for Plant Science, The University of Queensland, Warwick, Australia, Robert R. Klein, USDA-ARS, Southern Plains Agricultural Research Center, College station, TX and Ian Small, The University of Western Australia, Crawley, WA, Australia

To date, one of the most successful approaches to boosting seed yield of crop plants is by breeding hybrids and taking advantage of heterosis. Avoidance of self-pollination during hybrid seed production exploits mitochondrial genes that cause cytoplasmic male sterility (CMS). CMS is induced by mitochondrial genes that cause abortion of pollen development and the nuclear-encoded restorer of fertility (*Rf*) genes act by suppressing the expression of these CMS-specific gene products. The majority of Restorer-of-Fertility-like (RFL) proteins in cereals belong to a specific clade of the pentatricopeptide repeat (PPR) proteins family. PPRs are predicted to be located to organelles where they influence processing of specific transcripts. In this study we developed a bioinformatics pipeline to search for RFL-PPR sequences in genomic data sets of 44 sorghum accessions. This sequence knowledge was used to design a RFL-capture approach aimed at targeted-enrichment and identification of RFLs encoded in the genomes of three sorghum lines with different restoring capabilities. The capture approach and subsequent sequence analysis allowed the identification of homologous sequences of all 21 reference *RFL* genes in all three studied sorghum lines. Sequence analysis and comparison of genomic positions of the previously published candidates for major restorer genes in sorghum combined with our current understanding of how PPR proteins recognize their RNA targets allowed narrowing down the candidates for the *RF2* and *RF5* restorers to only single genes located on chromosome 2 and 5, respectively. Current studies are aimed at validation of the restoring capability of the sorghum RF2 and RF5 candidates.

P0884: Maize, Sorghum, Millet, Sugar Cane, and related

Evaluation for Fodder Quality in Sorghum RIL Population and Germplasm Reference Set in Contrasting Water Regimes

Vinutha K Somegowda[1], Kvsv Prasad[2], Ravi Devulapalli[2], AnilKumar Vemula[1], Abhishek Rathore[3], Michael Blummel[4] and **Santosh Deshpande**[1], (1)ICRISAT, Hyderabad, India, (2)ILRI, Hyderabad, India, (3)ICRISAT, Patancheru, Hyderabad, India, (4)ILRI, Addis Ababa, Ethiopia

Drought (midseason or terminal) is a regular and recurring event in arid and semi-arid regions of (approximately 30% total area) and are inhabited by 20% of the total world population. The reduction in crop production and yield caused by drought has direct effect on livelihood of farmers (and their families) that in turn affects the yield from livestock (draft capacity/milching). The current dry fodder production levels are 138 million tonnes against predicted demand of 526 million tonnes by 2020, which entails better emphasis on breeding objectives for improving fodder yield and quality, including grain. The terminal drought stress is known to reduce the grain yield and no severe effects on biomass yield; nonetheless no data is recorded on quality of the biomass (fodder). A 5% variation in the key fodder quality trait, *in vitro* organic matter digestibility (IVOMD) is associated with 20% of price premium. Hence to determine the effect of drought on fodder quality, a sorghum RIL population and reference set was field evaluated for fodder yield and quality under two water regimes (drought stress and control). The experimental design was alpha lattice with three and two replications for RIL population and germplasm reference set, respectively. The crop management practices were same except that irrigation was withheld 50 days after sowing under stress conditions. The data from agronomic, physiological and biomass subjected to NIRS (near infrared spectroscopy) to assess the fodder quality are presented.

P0885: Maize, Sorghum, Millet, Sugar Cane, and related

Effects Dynamics of Repulsion Linked QTLs during Sorghum Development Contributes to Plant Height Heterosis

Qi Mu, Anna Underhill, Xin Li, Xianran Li and Jianming Yu, Iowa State University, Ames, IA

Heterosis, the better performance of a hybrid than its parents, plays an important role in crop breeding. Pseudo-overdominance is one of four theories (dominance, overdominance, pseudo-overdominance, and epistasis) proposed to explain heterosis but with only one supporting case. Recently, we demonstrated the pseudo-overdominance heterosis of final plant height by identifying an additional QTL, *qHT7.1,* near the known plant height gene *Dw3* with sorghum recombinant inbred lines (RIL). Two QTLs, *qHT7.1* and *Dw3*, are 3Mb away and in repulsion phase between two parents. With two loci in repulsion phase between two inbreds, heterosis in the hybrid can appear as a single locus with an overdominance mode of inheritance – pseudo-overdominance. The high frequency of repulsion linked haplotype (57%) in sorghum diversity panel suggests the potential application of these two QTLs in breeding program. To better understand the effect dynamics of *qHT7.1* and *Dw3*, we measured the plant height of this RIL population across entire growing season for three consecutive years. The effects of *qHT7.1* and *Dw3* arise together and can be detected as earlier as the transition stage from vegetative to reproductive phase. *Dw3* reaches the highest effect at the booting stage, while the effect of *qHT7.1* keeps increasing until the final stage. Similar additive effects were detected for *qHT7.1* and *Dw3*, with *Dw3* has slightly higher values through the entire growing season. In addition to QTL mapping, identifying the gene underlying *qHT7.1* and further phenotypic analyses will shed light on the genetic basis and molecular mechanisms of plant height and heterosis.

P0886: Maize, Sorghum, Millet, Sugar Cane, and related

Quantitative Genetic Dissection of Seedling Chilling Tolerance in Sorghum

Sandeep R. Marla[1], Terry Joseph Felderhoff[1], Gloria Burow[2] and Geoffrey Morris[3], (1)Kansas State University, Manhattan, KS, (2)USDA-ARS, Lubbock, TX, (3)Department of Agronomy, Kansas State University, Manhattan, KS

Understanding the genetic basis of climate adaptation in diverse crops is critical for crop improvement. We utilize sorghum *(Sorghum bicolor* [L.] Moench), a tropical-origin C4 cereal crop, adapted to chilling in northern China to investigate its adaptation. The objective of our study is to identify loci conferring chilling tolerance in Chinese sorghum. We used US × Chinese bi-parental recombinant inbred lines (RILs) (n = 670) generated by crossing three chilling-tolerant Chinese accessions (Niu Sheng Zui, Kaoliang, and Hong Ke Zi) with chilling-sensitive US inbred BTx623. We phenotyped these 670 RILs for their response to early-season chilling stress in April (average temperatures below 10°C) in five environments, with two replicates per environment, in Kansas. Seedling vigor and cold damage ratings were determined based on their phenotypic appearance in response to chilling. We identified 43,320 single nucleotide polymorphisms using genotyping-by-sequencing that were used for joint linkage mapping (JLM). Using the seedling phenotypes and genomic data, JLM and linkage mapping identified genetic regions on chromosomes 4, 9, 3, and 7 associated with chilling tolerance. Chilling tolerance QTL identified on chromosomes 4 and 9 using JLM are within 1 cM distance from *Tannin1* and *Dwarf1*, respectively. In this study, using higher number of RILs with high-density markers enabled us to identify rare recombinants between chilling tolerance QTLs and undesirable traits (seed tannin and tall stature) present in Chinese sorghums. Identifying genes underlying sorghum early-season chilling tolerance can decipher mechanisms of chilling adaptation and can aid in developing seedling cold tolerance in tropical crops.

P0887: Maize, Sorghum, Millet, Sugar Cane, and related

Genetic Diversity and Selection Signatures in the Senegalese Sorghum Landraces

Jacques Martin Faye[1], Fanna Maina Assane Mamadou[2], Zhenbin Hu[3], Sandeep R. Marla[2], Daniel Fonceka[4], Ndiaga Cisse[5] and Geoffrey Morris[3], (1)Kansas State University, Department of Agronomy, Manhattan, KS, (2)Kansas State University, Manhattan, KS, (3)Department of Agronomy, Kansas State University, Manhattan, KS, (4)Centre d'Etudes Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), Thies, Senegal, (5)ISRA-Senegal, Bambey, Senegal

Improving crop adaptation to diverse environments is essential for food security in smallholder farming systems. Genomic signatures of local adaptation are recorded in crop genomes. In the West African country of Senegal, guinea sorghums are grown in humid regions, whereas durra sorghums are grown in dry areas. We tested whether climate adaptation is attributed to genomic regions associated with known water use-responsive and early maturity genes in durra, or known late maturity and panicle architecture genes in guinea. We constructed high-density single nucleotide polymorphisms (SNP) map from 421 Senegalese sorghum accessions available at the US Germplasm Resources Information Network. Population structure analysis showed genetic differentiation along the rainfall gradient, congruent with main agro-ecological zones. Genome-wide patterns of variation revealed positive selection along durra genomes. Because Ethiopia is known as the center of origin of durra, we tested the particularity of these positive selections across Ethiopian, West African, and Senegalese durra. Selection signatures on chromosome 7 were common to all African durra, those on chromosomes 5 and 10 were common to West African Sahelian durra. Sweeps on chromosome 2 was specific to Senegalese durra. Genes controlling flowering time, drought tolerance, and inflorescence architecture traits were found within selective sweeps. We also tested whether alleles associated with photoperiod sensitive and panicle architecture co-localize with known candidate genes. Association studies identified 178 SNPs for photoperiod sensitive and 48 SNPs for panicle compactness, located within or near predicted *a priori* candidate genes. These results can provide new resources for improving local adaptation of sorghum.

P0888: Maize, Sorghum, Millet, Sugar Cane, and related

Population Genomics of Sorghum (*Sorghum bicolor*) across Diverse Agroclimatic Zones of Niger

Fanna Maina Assane Mamadou[1], Sophie Bouchet[2], Sandeep R. Marla[1], Zhenbin Hu[3], Jianan Wang[3], Aissata Mamadou[4], Magagi Abdou[5], Abdoul-Aziz Saidou[6] and Geoffrey Morris[3], (1)Kansas State University, Manhattan, KS, (2)INRA, Clermont-Ferrand, France, (3)Department of Agronomy, Kansas State University, Manhattan, KS, (4)Institut National de la Recherche Agronomique du Niger, Niamey, Niger, (5)La Sahelienne des Semences, Maradi, Niger, (6)Université Dan Dicko Dankoulodo, Maradi, Niger

Improving adaptation of staple crops in developing countries is important to ensure food security. In the West African country of Niger, the staple crop sorghum (*Sorghum bicolor* L. Moench) is cultivated across diverse agroclimatic zones, but the genetic basis of local adaptation has not been described. The objectives of this study were to characterize the genomic diversity of sorghum from Niger and to identify genomic regions conferring local adaptation to agroclimatic zones and farmer preferences. We analyzed 516 Nigerien accessions for which local variety name, botanical race, and geographic origin were known. We discovered 144,299 single nucleotide polymorphisms (SNPs) using genotyping-by-sequencing (GBS). We performed discriminant analysis of principal components (DAPC), which identified six genetic groups, and performed a genome scan for loci with high discriminant loadings. The highest discriminant coefficients were on chromosome 9, near the sorghum ortholog of maize flowering time adaptation gene *Vgt1*. Next, we characterized differentiation among local varieties and used a genome scan of pairwise F_{ST} and Tajima's *D* values to identify SNPs associated with specific local varieties. Comparison of varieties named for light- versus dark-grain identified differentiation near *Tannin1*, the major gene responsible for grain tannins. These findings could facilitate genomics-assisted breeding of locally-adapted and farmer-preferred sorghum varieties for Niger.

P0889: Maize, Sorghum, Millet, Sugar Cane, and related

Newly Developed Melanaphis Resistant Sorghum Lines in Haiti Show a Strong Selective Sweep at a Locus on Chromosome 6 Collocated with the Known RMSE1 Gene

Jean Rigaud Charles[1], Kebede T Muleta[2], Geoffrey Morris[2] and **Gael Pressoir**[3], (1)Chibas - Quisqueya University, Port au Prince, Haiti, (2)Department of Agronomy, Kansas State University, Manhattan, KS, (3)Chibas - Quisqueya University, Bon Repos, Croix des Bouquets, Haiti

Infestation by a new biotype of aphids from the *Melanaphis* genus have occurred in Haiti in 2015, 2016 and 2017. These aphids never infested sorghum in Haiti before the year 2015 (this pest was only present on sugarcane). In 2016 over 80% of the sorghum production was lost because of this new pest.

After 2015, 252 new *Melanaphis sacchari* resistant sorghum lines have been developed by Chibas from its population undergoing recurrent selection for grain and stem sugar yields. These lines have been genotyped by genotyping by sequencing (GBS) along with a sorghum diversity panel including sorghum from all major races.

Based on the distribution of neutral SNPs we could detect a strong selective sweep using an Fst oulier method at a locus on the short arm of chromosome 6 collocated with the known RMSE1 Melanaphis resistance gene.

P0890: Maize, Sorghum, Millet, Sugar Cane, and related

QTL Mapping and Molecular Characterization of Sugarcane Aphid Resistance Gene in Sorghum

Jian Huang[1], Yinghua Huang[2], Michael Anderson[1] and Liuling Yan[1], (1)Oklahoma State University, Stillwater, OK, (2)USDA ARS, Stillwater, OK

Sorghum is a multipurpose drought-tolerant C4 plant, grown for food, feed, fiber and fuel. The sugarcane aphid (SCA), *Melanaphis sacchari* (Zehntner) is currently one of the most important insect pests of grain and forage sorghum in the U.S. The SCA can cause significant damage to plant and severe yield loss in sorghum. Host plant resistance to SCA is a sound method for the control of this devastating pest. Mapping genes for resistance to the SCA is an important aspect of understanding the molecular basis of host plant defense in sorghum. The objective of this study is to analyze the genetic control of resistance to SCA and locate resistance loci to specific genomic regions of sorghum. A mapping population was developed by crossing an elite line, BTx623 (susceptible parent) with a resistant donor (sugarcane aphid resistant line), then the

resulted F_2 population was genotyped using SSR markers. Simultaneously, all $F_{2:3}$ progenies were evaluated by screening those seedlings of the entire $F_{2:3}$ families against virulent SCA in growth chambers. With the collected two sets of data from genotyping and phenotyping experiments, linkage analysis was performed using the mapping software Map-Maker 3.0. QTL analysis was performed using Windows QTL Cartographer 2.5. In this way, the linkage relationship between SSR markers used in genotyping and SCA resistance loci was established, and then the resistance QTL was mapped to a specific location of the chromosomes of sorghum. In summary, SSR markers closely-linked to the resistance QTLs and identification of chromosomal regions responsible for sugarcane aphids contribute to both map-based cloning and marker-assisted breeding of new hybrids or cultivars with genetic resistance to sugarcane aphids.

P0891: Maize, Sorghum, Millet, Sugar Cane, and related

Genomic Selection in Sugarcane in Florida

Per Hilding McCord, USDA Agricultural Research Service, Canal Point, FL and Sushma Sood, USDA-ARS, CANAL POINT, FL

We are interested in testing the ability of genomic selection to predict the performance of new sugarcane clones with regard to important agronomic and disease resistance traits, as part of the breeding program at the USDA-ARS Sugarcane Field Station at Canal Point, Florida. For evaluation of the technique, we are phenotyping 416 individuals from an early stage (Stage 2) of the Canal Point program, and an additional 18 individuals from breeding programs in Louisiana. Yield, sucrose content, and disease data were measured on Stage 2 clones in an unreplicated trial in the 2015-2016 season. All 434 clones are being evaluated in a replicated trial for the 2017-2018 season, and have currently been phenotyped for stalk population, stalk diameter, and resistance to brown and orange rust via artificial inoculation. All individuals were genotyped via capture sequencing using 10,000 optimized probes, resulting in 21,277 markers. Genomic selection models have been developed using GBLUP. Prediction accuracy, or the correlation between predicted and measured values, varied from -0.009 for Brix in the Stage 2 trial, to 0.41 for stalk population in the full trial. Additional genomic selection models are being explored, such as the addition of large-effect single markers from genome-wide association analysis (GWAS) of the existing dataset, and the efficacy of a reduced set of markers is also under investigation.

P0892: Maize, Sorghum, Millet, Sugar Cane, and related

The SWEET Family Gene Evolution and Expression in *Saccharum* spp.

Weichang Hu[1], Xiuting Hua[2], Qing Zhang[3], Jianping Wang[4], Qiaochu Shen[5], Xingtan Zhang[1], Kai Wang[5], Qingyi Yu[6], Ray Ming[7] and **Jisen Zhang**[8], (1)Center for Genomics and Biotechnology, Fujian Provincial Key laboratory of Haixia applied plant systems biology, Haixia Institute of Science and Technology,Fujian Agriculture and Forestry University, Fuzhou, China, (2)Fujian Agriculture and Forestry University, China, (3)Center for Genomics and Biotechnology, Haixia Institute of Science and Technology (HIST)□Fujian Agriculture and Forestry University,, China, (4)Agronomy Department, University of Florida, Gainesville, FL, (5)Center for Genomics and Biotechnology, Haixia Institute of Science and Technology (HIST)□Fujian Agriculture and Forestry University, China, (6)Texas A&M AgriLife Research, Dallas, TX, (7)FAFU and UIUC-SIB Joint Center for Genomics and Biotechnology, Fuzhou, China, (8)Center for Genomics and Biotechnology, Fujian Provincial Key laboratory of Haixia applied plant systems biology, Haixia Institute of Science and Technology,Fujian Agriculture and Forestry University, Fujian, China

Sugarcane (*Saccharum spp*., Poaceae) contributed to ~80% total sugar and 40% of ethanol production in the world. As increasing sugar content is a major goal for sugarcane breeding, efforts to characterize the genes involved in sugar transport at the molecular level are growing significantly. The SWEET (Sugars Will Eventually be Exported Transporters) gene family is a newly identified group of sugar transporters, which play an indispensable role in sugar efflux, phloem loading, plant-pathogen interaction, nectar secretion and reproductive tissue development. However, little knowledge of SWEET are available for *Saccharum* due to lack of genome information for the crop comprising of complex genetic background.In this study, 22 SWEET genes were identified from *S. spontaneum BAC* libraries sequences. Ka/Ks analysis suggested that these genes except *SsSWEET4d* have experienced the purging of deleterious variations favored by strong purifying selection after the divergence between sorghum and S. spontaneum. Phylogenetic analyses of SWEETs from 11 representative plant species showed that the SWEETs could be divided into four clades. The gene expansions of SWEET family were mainly caused by the recent gene duplication in dicot plant, while, were attributed by the ancient whole genome duplication (WGD) in monocot plant species. The gene expression profiles were performed based on the RNA-seq analysis. *SWEET1a* was dominantly expressed in the transitional zone and maturing zone of the gradient leaf sections, while, *SWEET1b* was mainly expressed in the leaf tissues and the mature zone of the leaf in the two major *Saccharum* species, displayed expressional pattern with the peak at the end of night during the diurnal cycles. Both *SWEET2s* were more abundant in the sink-source transition zone than other zones of the leaf sections. *SWEET2b* was mainly expressed in the leaf of the two major *Saccharum* spp. Three *SWEET13s* showed similar expression patterns and were observed to be dramatically increased for their transcriptome from the maturing to mature zones. Of them, *SWEET13c* was extremely highly expressed in all the examined tissues. *SWEET13a* had a diel peak expression at the end of night period in *S. spontaneum*, but displayed no diurnal expression pattern in *S.officinarum*. *SWEET1a\2a\4a\4b\13a\16b* were also observed to be differential expressed between *S. officinarum* and *S. spontaneum*.Our study revealed the gene evolutionary history for the SWEET in *Saccarhum* and. *SWEET1b* was sucrose starvation-induced genes and involved in the sugar transportation in the high photosynthetic zones; *SWEET13c* was the key player in the efflux of sugar transportation in the mature photosynthesis tissues, *SWEET4a* and *SWEET4b* were mainly involved in the sugar transportation in the stalk. *SWEET1a\2a\4a\4b\13a\16b* were suggested to be the genes contributing to the differentia of sugar contents between *S. spontaneum* and *S. officinarnum*. Our results are valuable for further functional analysis of SWEET genes and utilization of the SWEET genes for technological improvement of *Saccarhum* for biofuel production.

P0893: Maize, Sorghum, Millet, Sugar Cane, and related

Transgenic Expression of Glyphosate Tolerant and Cane Borer Resistant Genes in Sugarcane (*Saccharum officinarum* L.)

Zahida Qamar[1,2], Idrees Ahmad Nasir[1], Bushra Rashid[1] and Tayyab Husnanin[1], (1)CENTER OF EXCELLENCE IN MOLECULAR BIOLOGY (CEMB), Lahore, Pakistan, (2)University of Toronto, Toronto, ON, Canada

Sugarcane (Saccharum officinarum L.) is an economically important cash crop. Quality yield is threatened by the damages of cane borers (Chilo Infuscatellus) and weeds. Two problems were addressed through the expression of modified two cane borer resistant genes CEMB-Cry1Ac, CEMB-Cry2A and glyphosate tolerant CEMB-GTGene. For higher expression, modified synthetic genes were designed according to the sugarcane genome by using the Codon optimization tool of Integrated DNA Technologies (IDT). One Variety was screened for further generation's level field study from SPF-213, SPF-234, HSF-240, and CPF-246, through tissue culturing response, glyphosate spray assays and transformation efficiency.
Double selection (Kanamycine (50mg/L) + glyphosate (50mM) gave 34% of SPF-213, 40% of SPF-234, 29% of HSF-240 and 81% of CPF-246 resistant calli. These transgenic shoots were confirmed through GUS assay and PCR analysis by using glyphosate gene-specific primers. The transgenic plants were treated with three doses, 900mL/acre, 1100mL/acre and 1200mL/acre of glyphosate sprays. The transformation efficiency 1.1% for SPF-213, 1.3 for SPF-234, 1% for HSF-240, and 1.5% for CPF-246 was observed. In V0, V1 and V2 generations, confirmation for stable integration of the transgenes was carried out by Sothern Blot, Dipstick assay and Enzyme-Linked Immunosorbent Assay (ELISA), Biotoxicity assays and glyphosate sprays.
It was concluded that with the expressions of CEMB-Cry1Ac, CEMB-Cry2A, and CEMB-GTGene, genes sugarcane variety CPF-246 efficiently resist against the cane borers (70-100% mortality) and was highly tolerant for glyphosate spray assay (1200mL/acre). These advanced lines can be very helpful for achieving higher and sustainable yields free of insect pests and weeds damages.

P0894: Maize, Sorghum, Millet, Sugar Cane, and related
Promoter Region Characterization of Co-Expressed Genes in Sugarcane
Augusto L. Diniz, Danielle I. R. Silva, Mauro de M. Oliveira, Felipe ten Caten and Glaucia M. Souza, Department of Biochemistry, Institute of Chemistry, University of Sao Paulo, Sao Paulo, Brazil
Gene co-expression network studies lead to the identification of biologically important patterns in transcriptomic data. A promising approach to study co-expression genes, and integrate the genome to transcriptome information is to characterize their promoter region and identify transcription factors (TFs) that may be involved in gene co-regulation. In this scenario, we aimed to identify TFs Biding Sites (TFBS) in specific co-expression profiles of two different microarray dataset, one from three sugarcane ancestral genotypes (*Saccharum officinarum*, *S. spontaneum* and *S. robustum*) in addition to the hybrid RB86-7515; and a second from field and greenhouse drought experiments, performed with the commercial hybrids SP80-3280, SP90-1638, RB86-7515, RB85-5536 and RB92-579. Co-expression modules were detected using the WGCNA R package and sequences up to 2000 nucleotides, upstream from the transcription start site of each gene, were investigated for TFBS discovery, through an in silico approach, using the Multiple Expectation Maximization for Motif Elicitation (MEME) algorithm. We identified modules correlated to important features of carbon partitioning and drought response. The most common motif found is responsible for anchoring TFs with AP2 / ERF domain. In addition, a particular TFBS architecture including motifs involved in the binding of TFs from MYB, DREB, NAC and SPL families to target genes was identified. This information allied to previous results indicating ABA as a major player on drought responses points to the importance of ABA-dependent and independent pathways crosstalk during water stress in sugarcane.

P0895: Maize, Sorghum, Millet, Sugar Cane, and related
Sugarcane EST Project (SUCEST) Re-Annotation for Improved Gene Expression Analysis
Felipe ten Caten, Danielle I. R. Silva, Augusto L. Diniz and Glaucia M. Souza, Department of Biochemistry, Institute of Chemistry, University of Sao Paulo, Sao Paulo, Brazil
Due to its large genome size and high ploidy sugarcane genomic and transcriptomic studies are challenging. The SUCEST (Sugarcane EST) Project was a prominent effort for characterization of sugarcane transcriptional landscape from high-quality ESTs, enabling the identification of 43,141 transcripts (Sugarcane Assembled Sequences - SAS). Studies based on this information led to the design of microarray experiments related to specific gene catalogues, such as signal transduction (SUCAST) and metabolism (SUCAMET), and also genes involved in general functions (CaneRegNet). Nevertheless, many sequences lack specific annotation. Most of the SAS represented in the CaneRegNet array for example, are unknown sequences or have similarity with hypothetical proteins, limiting the information available and requiring additional analysis steps for the investigation of transcripts of interest. To overcome these drawbacks and provide a more informative dataset we are conducting the re-annotation of the SUCEST sequences, applying up-to-date information available at different databases. Using Blast2GO pipeline and UniProtKB/Swiss-Prot, a manually curated protein database, we obtained a valid match for ~73% of SAS (e-value < 0.001), enabling a Gene Ontology (GO) association for 53.7% of sequences with 4,919 distinct terms. Most abundant categories include binding and catalytic activity as well as metabolic and cellular process. Regarding SAS included in CaneRegNet array, 87.2% of them match with annotated proteins and 72.5% are associated to GO terms. Next steps include the search against broader databases, the investigation of non-coding elements, association with the best matches against the sugarcane genome assembly and integration of results with tools available at http://sucest-fun.org platform.

P0896: Maize, Sorghum, Millet, Sugar Cane, and related
Identification *in silico* of the Promoter Region of Genes in Sugarcane Variety SP80-3280
Mauro de M. Oliveira[1], Ígor Bonadio[2], Alan mitchell Durham[2] and Glaucia M. Souza[1], (1)Department of Biochemistry, Institute of Chemistry, University of Sao Paulo, Sao Paulo, Brazil, (2)Department of Computer Science, Institute of Mathematics ans Statistics, University of Sao Paulo, Sao Paulo, São Paulo, Brazil
Genes in higher organisms generally present a regulatory region that is upstream of the coding region. The separation between the promoter region and the gene is marked by the transcription start site (TSS). However, the TSS is represented with only two nucleotides and is generally a degenerate sequence. Thus, the determination of the promoter and coding regions becomes a challenging task. Currently, biological experiments exist that allow to locate and to identify the TSS in the genome. However these experiments cannot be performed for all organisms that are studied, especially those that are not model organisms. Thus, the identification of the promoter region for many organisms is generally performed by *in silico* analysis. The our group developed a new predictor for the promoter regions of the sugarcane genes using using only information of the location and content of TSSs to create a conditional random field probabilistic model (CRF). This model was trained on 21000 sequences of experimentally validated arabidopsis promoter regions. We then applied this trained model in 18,000 sugarcane sequence

with validated TSSs. These sequences included 1000 nucleotides (nt) upstream the UTR region plus the 5' UTR region. Our predictions predicted 40% more TSSs within 50nt of the real one in 40% more sequences than the other prediction programs. These results indicate the robustness of our prediction model even when there is no experimental evidence from the promoter region of the target organism, or even when there is little information about the organism studied.

P0897: Maize, Sorghum, Millet, Sugar Cane, and related

Phenotypic Diversity and Heritability of Major Traits in Eragrostis Tef in Israel

Shiran Ben Zeev and Yehoshua Sarnaga, Hebrew University of Jerusalem, Rehovot, Israel

Tef [*Eragrostis tef* (Zucc.) Trotter] is a C4 grain cereal, commonly grown in Ethiopia, where it was presumably domesticated. The worldwide interest in tef cultivation and consumption has considerably increased during the last decades due to its nutritional potential and being a gluten free grain. In addition, big immigration waves of Ethiopian Jews increased the demand for tef grain in Israel. Thus, interest in tef was renewed in Israel since it was first examined in the 1930's as a forage crop.

Over 400 tef lines held in the Israeli gene bank were phenotyped under common garden (screen house) conditions in 2015. A subset of 270 lines, representing the collection's phenotypic diversity, was selected and evaluated under field condition in the Israeli Negev area in 2016. A third trial was conducted in central Israel in 2017 in single plant plots (to reduce admixtures). Seven traits where evaluated in all three years: days to flowering, plant height, spike length and shape, biomass weight, grain yield, and 1000 grain weight. The collection exhibited a wide diversity for the measured traits under all three environments. Broad sense heritability estimates, calculated for each pair of years and all three years, varied between 0 and 0.7, with days to flowering, 1000 seed weight and plant height having the highest values.

This study demonstrated that tef can successfully grow and produce under Mediterranean conditions. The tested lines are currently subjected to marker detection analysis, as a basis for future breeding.

P0898: Maize, Sorghum, Millet, Sugar Cane, and related

Functional Genomic Analysis of Abiotic Stress Tolerance in Pearl Millet

Sudhakar Reddy Palakolanu[1], **V B Reddy Lachagari**[2], Sivasakthi Kaliamoorthy[1], Richa Yashwekar[1], Rahul Nitnavare[1], Pooja Bhatnagar Mathur[1], Kiran Sharma[1], Vincent Vadez[1], Boney Kuriakose[3], Sivarama Prasad Lekkala[2], Navajeet Chakravartty[2] and Krishna Mohan[2], (1)ICRISAT, Hyderabad, India, (2)AgriGenome Labs Pvt Ltd, Hyderabad, India, (3)AgriGenome Labs Pvt Ltd, Kochi, India

Pearl millet [*Pennisetum glaucum* (L.) R. Br.], is an important crop of the semi-arid and arid regions of the world. It is capable of growing in harsh and marginal environments with the highest degree of tolerance to drought and heat stresses among cereals with diversity in germplasm showing significant variation in phenotype with response to abiotic stress factors. This makes the pearl millet a unique model among other cereals to study abiotic stress factors. We have investigated transcriptome dynamics of pearl millet under vapour pressure deficit (VPD) to study water saving mechanism in drought tolerant and susceptible lines of pearl millet through RNASeq and miRNA sequencing at various conditions. We evaluated the physiological performance, Transcriptome and microRNA expression profiles of two pearl millet genotypes, ICMR 1122 and ICMR 1152, showing contrasting response to high VPD conditions. A total of 158 known and 10 novel miRNAs were identified from the miRNA profiling of which significantly differentially expressed miRNAs and their targets were studies for their association in drought tolerance mechanism along with their transcriptome profiles in root and leaf samples. A total of 2,724 DEGs were identified under drought treatment from the transcriptome analysis. These genes comprise 289 Transcription Factors (TFs) representing Basic Helix-loop Helix (bHLH), Ethylene Response Factors (ERFs), myeloblastosis (MYB), No apical meristem (NAC), and WRKY amino acid motif (WRKY) type major families known to be involved in the mechanism of stress tolerance.

P0899: Maize, Sorghum, Millet, Sugar Cane, and related

Unlocking Crop Genomic Diversity for Breeding: Lesson from the Pearl Millet Genome

Yves P. Vigouroux, IRD, Montpellier, France

lant breeding uses more and more genomic diversity to accelerate the development of new varieties. Several new approaches have been developed to identify functional diversity variation using such genomic data. Past evolutionary history of the crop shapes both its neutral and functional diversity. Moreover, wild diversity, still largely unexplored, could also now be more efficiently studied and used. Finally, new phenotypes could be used for breeding from root architecture to variety shaped bacterial community. We will illustrated these different strategies and studies in an important Sahelian cereal, pearl millet. Using whole genome sequence and new breeding traits, such advances could considerably enhance our ability to breed better crop in a hotter climate.

P0900: Maize, Sorghum, Millet, Sugar Cane, and related

Accelerating Gene Discovery in Setaria viridis to Study C4 Photosynthesis: From Fine Mapping to Community Resources

Hui Jiang[1], Pu Huang[1], Kerrie W. Barry[2], Jerry Jenkins[3], Jane Grimwood[3], Jeremy Schmutz[3], Xiaoping Li[1], Max Feldman[4], Ivan Baxter[5], Elizabeth A. Kellogg[1] and Thomas P. Brutnell[6], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (4)Donald Danforth Plant Sciences Center, St. Louis, MO, (5)USDA-ARS/Donald Danforth Plant Science Center, St. Louis, MO, (6)Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO

Setaria viridis is an excellent model system for the study of C_4 grasses due to its small stature, short life cycle, small genome size (~515Mb), ease of crossing and capacity for transformation. To accelerate gene discovery in *Setaria viridis*, we have developed an NMU-mutagenized population consisting of 20,000 M2 families. Sixty one sequenced mutants showed a median of 66 homozygous nonsynonymous mutations per family. Bulked Segregant Analysis followed by deep sequencing (BSADeepSeq) was used to rapidly identify several putative candidate genes underlying traits of interest including a sparse panicle mutant (Nat Plants. 2017 Apr 18;3:17054), and a gene putatively required for Vitamin B6 biosynthesis that conditions a mottled pale green phenotype when disrupted. To identify genes that may function in the carbon concentrating pathway associated with C_4 photosynthesis, approximately 3,000 M2 families were screened under ambient CO_2 conditions and

~300 families were identified with pale green mutant phenotypes. Additional screening under low CO_2 conditions (~150ppm) identified a necrotic leaf (*nl1*) mutant that can be rescued under high CO_2 (~3,000ppm). BSADeepSeq enabled us to fine map the *nl1* phenotype to a 1Mb region. Fine mapping to identify the causative lesion is underway. These findings demonstrate the value of *Setaria viridis* as a tractable genetic model for rapid gene identification. We have also assembled a collection of ~650 diverse *Setaria* accessions collected throughout the US and Canada. These accessions have been sequenced at JGI-DOE and a subset has been propagated based on genome structural analysis as a foundation for GWAS. These seeds will be propagated and sent to USDA GRIN for seed distribution. We have also initiated the construction of seven recombinant inbred line (RIL) populations. Collectively, these genetic resources will help drive future genetic studies in Setaria.

P0901: Maize, Sorghum, Millet, Sugar Cane, and related

Unlocking the Genetic Resources of a Neglected and Underutilized Crop Species: A Case Study of Acha (*Digitaria* spp. (Kippist) Stapf)

Oluwatoyin Ogundipe, University of Lagos, Lagos, Nigeria

Acha is a crop endowed with wealth of nutritional and medicinal uses, yet the rate at which the crop is being lost is at alarming rate. Hence, the need for germplasm collection, preservation and utilization for future generations and crop improvement. This study carried out exploration trip on Acha collections (*Digitaria exilis* and *Digitaria iburua*) with a view to unlocking and investigating its genetic resources for the benefit of mankind. Five Northern states in Nigeria where Acha was cultivated were fully explored for sample collections. Germplasm collections was preceded by initial discussions with farmers on their fields, at market places and homes. At the end of the exploration tour, a total of thirty eight villages were visited out of which seventy five Acha collections were identified. The germplasm collected includes seed saved from previous harvest, purchased from the market, and those shared among colleagues. At the end of the trip, a pre project workshop was organized where farmers from the five states were invited and discussions on the uses, cultivation, processing and benefit of Acha was fully discussed. The prime focus of this workshop was to bring back Acha not only to the Northern Nigeria but to other parts of Nigeria to grow.

Keywords: Acha, Neglected, Genetic, Germplasm, *Digitaria spp.*

P0902: Maize, Sorghum, Millet, Sugar Cane, and related

High Density Genetic Map Construction and Comparative Genomics Analysis in St. Augustinegrass

Xingwang Yu and Susana R Milla-Lewis, North Carolina State University, Raleigh, NC

St. Augustinegrass (*Stenotaphrum secundatum*) is a warm-season grass species commonly utilized as turf in the southeastern US. However, molecular breeding on St. Augustinegrass has been slow due to limited genomic information. We constructed a high-density linkage map from "Raleigh" x "Seville" F1 mapping population using genotyping-by-sequencing approach. A total of 2,871 single-nucleotide polymorphisms (SNP) and 85 SSR markers were mapped on 9 linkage groups. The genetic linkage map had a total length of 1244.25 cM, with an average of 0.42 cM between markers. Comparative genomics analysis was carried out between St. Augustinegrass and other model grass species including foxtail millet, sorghum and rice. Widespread synteny and colinearity of genetic loci were found between the genomes of these species. In addition, several inter-chromosomal rearrangements, which differentiated St. Augustinegrass from other grasses, were identified in LG2, LG3, LG7, LG9. The high density genetic map and comparative genomics study provides a base for further functional gene cloning and genome sequence assembly, which will significantly benefit breeding efforts in St. Augustinegrass.

P0903: Maize, Sorghum, Millet, Sugar Cane, and related

The Trait Components That Constitute Whole Plant Water Use Efficiency Are Defined By Unique, Environmentally Responsive Genetic Signatures in the Model C4 Grass Setaria

Max J Feldman[1], Patrick Ellsworth[2], Noah Fahlgren[3], Malia Gehan[4], Asaph B. Cousins[2] and Ivan Baxter[5], (1)Donald Danforth Center Plant Science Center, St. Louis, MO, (2)Washington State University, Pullman, WA, (3)Bioinformatics Core Facility, Donald Danforth Plant Science Center, St. Louis, MO, (4)Donald Danforth Plant Sciences Center, St. Louis, MO, (5)USDA-ARS, Danforth Plant Science Center, St. Louis, MO

The complex relationship between plant growth and water use is largely determined by genetic factors that influence both the morphological and biochemical characteristics of plants. Improving the efficiency by which plants utilize water is an important agronomic breeding objective that can be translated to improve productivity in agriculture while simultaneously making it a more sustainable endeavor. To assess the genetic basis of water use efficiency (WUE) and trait plasticity, we have utilized a high-throughput phenotyping platform to quantify plant size and evapotranspiration of an interspecific *Setaria italica* x *Setaria viridis* recombinant inbred line population in both a well-watered and water-limited environment. Our findings indicate that measurements of plant size and water use in this population are strongly correlated and that using a linear modeling approach to partition this relationship into both the predicted values of plant size given water use and deviations from this relationship at the genotype level provides a useful framework to understand plant WUE. Biparental linkage mapping successfully identified several major pleiotropic loci that exhibit medium-to-large effects on most traits in addition to many smaller effect loci associated with fewer traits or specific to well-watered or water-limited environments. This study is the first report characterizing the genetic architecture of WUE in the model C_4 species *Setaria* and indicates that alleles derived from both wild and domesticated species can be utilized to predictably modulate trait values given a specified precipitation regime.

P0904: Maize, Sorghum, Millet, Sugar Cane, and related

Physiology and Complex Genetic Variation of Anaerobic Germination in Rice

Lina Castano-Duque, Duke University, Durham, NC and Sharmistha Ghosal, Fergie Quilloy, Shalabh Dixit, Thomas Mitchell-Olds

To achieve lower production costs for rice the current agricultural trend is to shift from transplanting seedlings to direct sowing of seeds. Those seeds may have to germinate under anaerobic conditions due to flash flooding. We are studying the physiology and genetic variation of anaerobic germination (AG) in rice natural populations, by using new analytical methods to integrate diverse functional information to identify complex trait loci in genome wide association studies (GWAS). In the first screening experiment we quantified AG% in 2,700 sequenced rice

lines after 7, 14 and 21 days of sowing in flooded and control conditions at International Research Rice institute and selected 120 lines with medium and high AG% to do a second screening. In the second screening we characterized AG%, root architecture, and shoot to root biomass ratios. We are currently using the characterized traits from the second screening to run GWAS and to select among the 120 lines the ones that will be used as parental lines to generate F3 families by single seed descent. Parental and F3 lines will be used to understand the physiology and genetics of the AG trait by characterizing their genetic, physiology and biochemical responses to different levels of anaerobic distribution treatment. We aim to identify candidate genes that influence quantitative variation for AG in rice, and determine their mode of action. The study of complex genetic variation and physiology in rice under anaerobic germination is necessary to produce seeds with traits that can fulfill societal and agronomical needs.

P0905: Oilseeds, Sunflower, and related

A Large-Scale Phenotypic and Transcriptomic Analyses of Domesticated Lines of Sunflower Under Flooding Stress

Joon Seon Lee, Lexuan Gao and Loren H. Rieseberg, University of British Columbia, Vancouver, BC, Canada

Understanding of how plants respond and adapt to different abiotic stresses is essential in crop science. The combination of detailed phenotypic analysis and the transcriptomic analysis under abiotic stress is a powerful tool for finding sets of genes that control a plant's response to abiotic stress. We have characterized phenotypic traits of 288 sunflower inbred lines from a cultivated sunflower association mapping (SAM) population under the flood stress. We measured multiple traits for possible flooding responses, such as biomass and chlorosis reduction and hypocotyl hypertrophy. We chose two inbred lines (the most resistant and the least resistant to the flood stress) for ultra high-throughput transcriptome sequencing. We have identified differentially expressed genes that potentially regulate the flood stress response in cultivated sunflowers and differentially spliced genes that may govern the expression patterns via diverse isoforms. The co-expression network of these stress-responsive genes highlights functional modules that may be used in flood stress response more broadly.

P0906: Oilseeds, Sunflower, and related

Sunflower Domestication in Space and Time

Nathan Wales[1], Melis Akman[1], Ray H. B. Watson[1,2], Sariel Hubner[3], Fátima Sánchez Barreiro[4], Bruce D. Smith[5], Kristen J. Gremillion[6], Loren H. Rieseberg[7], M Thomas P Gilbert[4] and **Benjamin K. Blackman**[1,2], (1)University of California, Berkeley, Berkeley, CA, (2)University of Virginia, Charlottesville, VA, (3)Galilee Research Institute, Kiryat Shmona, Israel, (4)Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark, (5)The Santa Fe Institute, Santa Fe, NM, (6)Ohio State University, Columbus, OH, (7)University of British Columbia, Vancouver, BC, Canada

Native American farmers living ~4000-5000 years ago transformed the common sunflower, *Helianthus annuus*, from a highly branched wild plant with small disks and small seeds into a staple oilseed crop that sports a single large head with large seeds on an unbranched stalk. We are pursuing three complementary approaches to understand how this process unfolded. First, we are conducting population genomic scans for selective sweeps to highlight candidate genes that may have contributed to the domestication process. Second, we have generated comparative transcriptomic data for an array of tissues and developmental stages in cultivated and wild sunflower, and tissue-specific patterns of candidate gene expression are facilitating association of candidate genes to domestication traits. Finally, we have assembled an archaeological time series of disks and achenes for direct dating and DNA sequencing. These samples date from ~3500 to ~400 years before present, and we have obtained useful endogenous DNA content from samples throughout this interval. Initial shotgun sequencing of DNA from these samples and from ethnographic collections from the historic period confirm an Eastern North American origin of domesticated sunflower, reveal an early genetic bottleneck, and highlight the further loss of landrace genetic diversity with the advent of modern oilseed production. In the future, we will integrate the results of our three approaches. By performing targeted sequencing of candidate genes from our archaeological DNA libraries, we will characterize the timing and order of selective sweeps and thus gain understanding into how the sunflower domestication syndrome was assembled through time.

P0907: Oilseeds, Sunflower, and related

Optiarch: Optimization of Plant Architecture in Sunflower for Yield Increase

Yves Sprycha[1], Constantin Jansen[2], Christian Harnack[1], Lisa Popiolkowski[1], Aleksandra Dimitrijevic[3] and **Renate Horn**[1], (1)University of Rostock, Rostock, Germany, (2)Strube Research GmbH & Co. KG, Söllingen, Germany, (3)Institute of Field and Vegetable Crops, Novi Sad, Serbia

The aim of the project is to optimize the plant architecture in sunflower by creating a more compact and efficient ideotype. We focus on narrow leaf and branch angles as well as on reduced height for increasing yield through higher plant densities. Candidate genes for our approach to optimize the plant architecture in sunflower are genes of the GA-signalling pathway and two genes, involved in leaf and branch angle development. A sunflower association panel of 384 accessions was used for scoring the traits. We selected 64 defined growth types for further association studies. The candidate genes will be analysed by amplicon targeted sequencing to identify SNPs or InDels that are associated with narrow leaf and branch angles as well as reduced plant height. Based on a SNP described for the *HaDELLA1* gene we have already designed an HRM marker to test the association panel for dwarfism caused by this mutation. The accessions were further divided into restorer and maintainer lines by application of two markers, one specific for the PET1 cytoplasm and one for the restorer gene *Rf1*. This allows a better planning of test crosses for hybrid performance. SNP-based markers will be developed to assist the introduction of interesting traits of plant architecture into existing breeding programs in sunflower. Performance trials using different plant densities will enable the identification of gene variations relevant for high yields per hectare. The project will offer the possibility to analyse genes with significant influence on plant architecture in sunflower.

P0908: Oilseeds, Sunflower, and related

Integrating Phenotypic and Genetic Diversity for Verticillium Wilt Resistance Breeding in Sunflower

Juan Francisco Montecchia[1], Carla Filippi[2], Monica Fass[2], Carla Maringolo[3], Ignacio Cerrudo[3], Matias Dominguez[4], Esteban Hopp[5], Alberto Escande[6], Julio Gonzalez[7], Julio DiRienzo[8], Daniel Avarez[9], Facundo Quiroz[6], Veronica Lia[2], Norma Paniego[5]

and Ruth Heinz[5], (1)Instituto de Biotecnología - Instituto Nacional de Tecnología Agropecuaria (INTA), Hurlingham, Buenos Aires, Argentina, (2)CONICET, Hurlingham, Argentina, (3)CONICET, Balcarce, Argentina, (4)INTA-CONICET, Pergamino, Argentina, (5)INTA, Hurlingham, Argentina, (6)INTA, Balcarce, Argentina, (7)INTA, Pergamino, Argentina, (8)Universidad Nacional de Córdoba,, Córdoba, Argentina, (9)INTA, Manfredi, Argentina

Fungal diseases represent a significant constraint for sunflower grain and oil yield production worldwide. One preponderant disease affecting sunflower crop is Verticillium Wilt (SVW), which is caused by the soil-borne fungal pathogen *Verticillium dahliae* (*Kleb.*). In Argentina, SVM is endemic and affects an area of 1.2 million hectares, where 70% of the sunflower crop is grown. Fungicide control is not effective against SVW. Thus, resistance breeding is a key strategy to cope with this affection.

In this work, we explore the genetic diversity and the phenotypic response against SVW of 164 inbred lines from an association mapping population (AMP) as the first step for a Genome-Wide Association Study (GWAS) approach to assist breeding for disease resistance. The main goals of this work were (a) to evaluate the response of the AMP to SVW by multivariate analysis, and (b) to assess the relationship between resistance response and genetic structure.

The phenotypic responses to SVW were evaluated in the AMP inbred lines during three consecutive seasons. Disease incidence, severity, intensity and AUDPC adjusted means were obtained using Generalized Linear Mixed Models. In parallel, the population sub-structure, and genetic diversity were estimated using a 18161 SNP-imputed loci matrix.Principal Component Analysis was conducted on the inbred´s phenotypic means for each variable scored, partitioning the data by different criteria (i.e., country of origin) to establish their relationship with the distribution of phenotypic variation.

Through this study we have found resistance sources across the AMP sub-structuring groups, supporting the implementation of GWAS to assist breeding on SVW-resistance.

P0909: Oilseeds, Sunflower, and related

Meta-Analysis of GWAS in Canola Blackleg (*Lepthosphaeria maculans*) Disease Traits Demonstrates Increased Power from Imputed Whole-Genome Sequence

Mulusew Fikere[1,2], Denise Barbulescu[3], Michelle M. Malmberg[1,2], German C. Spangenberg[1,2], Noel O.I. Cogan[1,2] and Hans D. Daetwyler[2,4], (1)Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Australia, (2)School of Applied Systems Biology, La Trobe University, Bundoora, Australia, (3)Agriculture Victoria, Grains Innovation Park, Horsham, Australia, (4)Agriculture Victoria, Bundoora, Australia

Blackleg disease resistance is a target trait for global canola (*Brassica napus* L.) improvement programs to minimize yield loss. To identify resistance genes and genomic regions, genome-wide association analysis of 585 canola diversity panels was performed using imputed whole-genome sequence (WGS) and lower density genotype-by-sequencing (GBS). Stringent quality control thresholds were applied on WGS and GBS variants. Eagle + Minimac was used to impute WGS. Genome-wide association study (GWAS) were performed using Effective Mixed Model Association eXpedited (EMMAX) in hardcoded (0,1,2) and dosage genotype coding formats. WGS GWAS with 1,234,708 million SNP detected a larger number of significant SNP (and achieved a lower false discovery rate; FDR) than GBS with 64,072 SNP. WGS genotypes in dosage format resulted in lower FDR than hardcoded genotypes. We identified several quantitative trait loci (QTL) for blackleg disease resistance. Combining blackleg traits in a meta-analysis resulted in increased detection power than single-trait analysis. The meta-analysis GWAS identified 27 significant ($P<0.0001$) regions for blackleg traits (survival rate and internal infection) on chromosomes A02, A05, C03 and A08, with the majority of significant SNPs on A02 followed by A08. The significant SNPs identified on A02 were all located in a 3685-kb region. This study provides insight into the genetic architecture and potential molecular mechanisms underlying canola *L. maculans* resistance.

P0910: Oilseeds, Sunflower, and related

Genetic Diversity for Yield and Quality Traits in Sesame

Naama Teboul, Faculty of Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

Sesame (*Sesamum indicum* L.) is an important oil-crop worldwide. Sesame has been evaluated as a health food. The seeds of sesame are used for cooking and baking, or crushed for producing high-grade edible oil or oily paste (Tahini) and sweets. In spite of its economic and social importance, sesame is an 'orphan crop' with limited research conducted so far.

Micronutrient malnutrition afflicts over two billion people worldwide and prevalent mostly in developing countries where the diet lacks diversity and based mainly on a single staple food. Therefore, enhancement of grain nutrients (biofortification), either agronomically (application of mineral fertilizers) or genetically (breeding), is considered the most promising and cost-effective approach to alleviating malnutrition and related health problems. In the current study a large collection of worldwide sesame genotypes was characterized for yield and quality traits. A wide genetic diversity in morpho-phisiological and micronutrients content was found among genotypes. An F_2 mapping population, derive from a cross between two contrasting genotypes, was used for genetic dissection of seed mineral-nutrient. Using high-throughput genotyping, we were able to identify and characterize novel QTLs conferring high seed nutrient contents. Our results demonstrate the potential of using sesame as a model for study of biofortification. Revealing the molecular basis of nutrient accumulation in sesame will contribute to the development of toolkits for high yielding sesame varieties with enhanced nutritional values.

P0911: Oilseeds, Sunflower, and related

High-Quality Genome Sequences of Cultivated (*Linum usitatissimum*) and Wild (*L. bienne*) Flax

Frank M. You[1], **Khalid Rashid**[1], Zhen Yao[2], Sylvie Cloutier[3], Wen Chen[3], Ming-Cheng Luo[4] and Tingting Zhu[4], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)AAFC, Morden, MB, Canada, (3)AAFC, Ottawa, ON, Canada, (4)University of California, Davis, CA

Khalid Y. Rashid: Agriculture and Agri-Food Canada

The cultivated flax (*L. usitatissimum*) and its wild progenitor (*L. bienne*) are important germplasm for flax breeding. Both species are diploids and have a haploid chromosome number of 15 with a genome size of ~370 Mb. The cultivated flax cv. CDC Bethune has been sequenced using Illumina platform (Wang et al. 2012) and sorted onto chromosomes (You et al. 2016). Additional sequencing using the PacBio Sequel system

was performed to ~60× to fill gaps and lengthen scaffolds. We also sequenced the wild flax accession LIN1917 using the long reads of the PacBio RS II system to ~50×. *De novo* assembly was performed using the canu software followed by polishing using Quiver or Arrow (for raw PacBio reads) and Pilon (for Illumina reads). The CDC Bethune assembly contained 2315 contigs in 359 Mb with an N50 of 2.13 Mb while LIN1917 assembled into 1152 contigs totalling 313Mb with an N50 of 0.90 Mb. To validate, orient and order the PacBio contigs to chromosomes, we constructed BioNano optical maps for CDC Bethune, Macbeth (pedigree similar to CDC Bethune) and LIN1917. With BioNano optical maps, consensus genetic map (Cloutier et al. 2012), BAC fingerprint-based physical map (Ragupathy et al. 2011) and the previous version of cultivated flax pseudomolecules, we generated high-quality, chromosome-based pseudomolecules with fewer and smaller gaps for both cultivated and wild flax. These pseudomolecules will be useful to study genome evolution, for genome-wide SNP discovery, QTL identification, and comparative genome analysis and to capitalize upon wild flax germplasm in breeding.

P0912: Oilseeds, Sunflower, and related

Creation of Short Internodes Overexpressed Jatropha Plant

Sompid Samipak and Siwapech Sillapaprayoon, Kasetsart University, Bangkok, Thailand

Jatropha curcas L. or physic nut is a biofuel plant that can grow widely in the tropical and subtropical areas. The oil squeezed out from seeds can be directly used in low power machine. Besides yield, convenience in harvesting is a major concern for farmers and company especially in industrial scale plantation. *J. curcas* plant often reaches 1.5-2 meter in height which causes an inconvenience to harvest the fruit. *Short Internodes* (*SHI*) gene in *J. curcas* has been identified and characterized. DNA construct containing *JcSHI* under the control of CaMV 35S promoter was created and transformed into *J. curcas* cotyledon by Agrobacterium co-cultivation method. Through tissue culture selection, thirty seven regenerated shoots were obtained. PCR using a primer pair specific for the 35S:JcSHI region (35S+R) was done to confirm the success of transformation, the correct amplified product was only found in plant No. 2, 11, 16, 17, 28, 29, 30, 31, 33-44 and 46. Semiquantitative RT-PCR was accomplished to analyze expression levels of *SHI* gene in the transgenic plants and found increase in expression levels in transgenic plant No.11, 16, 31, 39 and 44. Height of plants were measured in the tissue culture bottles and found no significant difference between the control group and the *JcSHI* overexpressed group though some plants showed stunted growth and short internode length.

P0913: Rice

Association Analysis using USDA Diverse Rice (*Oryza sativa* L.) Germplasm Collections to Identify Loci Influencing Grain Quality Traits

Jeremy D. Edwards, Trevis D. Huggins, Ming-Hsuan Chen, Aaron K. Jackson and Anna M. McClung, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR

The USDA rice (*Oryza sativa* L.) core subset (RCS) was assembled to represent the genetic diversity of the entire USDA-ARS National Small Grains Collection and consists of 1,794 accessions from 114 countries. The USDA rice mini-core (MC) is a subset of 217 accessions from the RCS and was selected to maximize genotypic, phenotypic and geographical diversity in a panel with a more manageable size for phenotyping and genotyping experiments. Recently, next generation resequencing data have become available for the MC accessions and high density fixed array SNP data have become available for 189 RCS accessions not included in the MC. In the current genome wide association study (GWAS), we use the new high density genotypic data along with previously reported and new phenotype data for grain traits, including grain dimensions, apparent amylose content (AAC), alkali spreading value (ASV, an indicator of gelatinization temperature), protein content, chalk percent, and agronomic traits that may influence grain quality, including plant height, and days to heading. The results identified several known starch genes, e.g., *Waxy* gene (*granule bound starch synthase 1* (*GBSS1*)), along with 11 novel loci for grain starch traits and 7 for grain chalk. Many loci were significantly associated with multiple grain quality traits suggesting mediated pleiotropic effects. Based on these results, molecular markers will be designed and deployed for use in marker assisted selection to produce new rice varieties with high grain quality.

P0914: Rice

G x E in the Rice Global MAGIC Population under Seedling Stage Drought

Brook T. Moyers[1], John McKay[1] and Amelia Henry[2], (1)Colorado State University, Fort Collins, CO, (2)International Rice Research Institute, Los Banos, Laguna, Philippines

Rice is a water-intensive crop and the primary calorie source for many developing countries. The challenge of increasing rice yields must be met under the additional constraints of a growing human population and changing climate. Part of the problem is the length of the modern breeding cycle, and accelerating this cycle means increasing the rate and ease of trait selection. To do this requires thorough knowledge of the causes of trait variation, both genetic and environmental. The rice Global MAGIC population is a new, highly recombinant multi-parent population developed at the International Rice Research Institute (IRRI). The Global MAGIC population has 16 parents, eight from each of the two major rice subspecies (*indica* and *japonica*), and exhibits major transgressive segregation for most phenotypes we have examined. In this population, we examine the genetic architectures of establishment, root, and yield component traits following seedling stage drought stress across three growing seasons, and compare these architectures across seasons and against a well-watered experiment with the same population. In leveraging this population, our research simultaneously identifies genomic regions underlying key traits and specific recombinant lines for input into breeding programs, directly benefiting IRRI's mission to improve food security and farmers' livelihoods.

P0915: Rice

MutTILL: Development of Induced Mutant Resources in Rice (*Oryza sativa* L.) and Genome-Wide Sequencing of Mutant Lines Using Next Generation Sequencing Technologies

V B Reddy Lachagari, Reddiah Bodhanapu, Sivarama Prasad Lekkala, Krishna Lalam, Navitha Kokkonda, Navajeet Chakravartty and Krishna Mohan, AgriGenome Labs Pvt Ltd, Hyderabad, India

Mutagenesis based TILLING (MutTILL) is a reverse genetic approach that can be applied in plant research and breeding programs to broaden the genetic variability and for the identification of novel mutant alleles in specific genes of interest with improved varieties. MutTILL method

is non-transgenic approach, further removing the barriers in marketing new varieties. It has been widely adapted to different species across genera to study the gene function. We have developed a large mutagenized and re-mutagenized populations, both of which are at different stages of development. Among these two populations, we have successfully identified few mutants i.e. plant height and early flowering using forward screening and the lines showing contrasting phenotype used for whole genome sequencing for identification of genes responsible for that particular trait. ~10 GB of 2x150 bp paired end data was generated with approx 25X coverage using Illumina Hiseq X. The SNPs/INDELs were identified between control and mutant lines and variants were annotated using MSU7 assembly. We believe that these developed mutant populations will help in the identification of novel mutations for various biotic and abiotic stresses in Rice.

P0916: Rice

Improving Rice Genome Annotation Based on Public RNA-Seq Data

Lili Hao, BIG Data Center, Beijing, China

Rice is one of the most important staple food for a large portion of the world's population and also a key model organism for cereal crops due to its great agricultural importance. In order to take maximum advantage of reference genome in extensive post-genomics studies, it is advisable to keep annotation information updating all the time by integrating the latest massive multiple-omics data. Here, we present a revision of the *Oryza Sativa Japonica* genome annotation: *BIGD_IC4R-1.0,* based on more than 500 recent available high-quality RNA-seq data sets along with annotation contributions from NCBI, EBI and UniProt, thereby providing substantive improvements over the previous version *MSU Rice Genome Annotation Project Release 7.0*. Our near-final release of the *BIGD_IC4R-1.0* is consisted of 57,905 protein coding genes, including 2,259 novel loci which do not overlap with the previous annotations. A total of 67 percent of these gene models are corroborated by evidence of expression and the structural annotation of over 20,682 loci in the *MSU7.0* has been updated. Moreover, the number of genes in *BIGD_IC4R-1.0* with splice variants is significantly increased compared with *MSU7.0*. In addition, 11,841 long ncRNAs were predicted from 658,655 StringTie and PASA assembled transcripts and then included in *BIGD_IC4R-1.0*. This updated genome re-annotation has revised hundreds of incorrect predicted gene models in rice and provided a number of alternatively spliced isoforms as well as long ncRNAs, thus would hopefully promote the future functional genomics and transcriptomics studies in *Oryza* plants.

P0917: Rice

Unravelling the Mosaic Structure of Rice Genomic Diversity.

Joao Garcia Nolasco Dourado Santos, Centre International de Recherche Agronomique pour le Developement, Montpellier, France

Since its domestication in Asia, rice has evolved a range of diverse cultivars adapted to multiple environments and regional preferences. This diversity benefited from broad and multiple foundation events which, in the context of the complex system of exchanges and migrations imposed by Man, allowed the further generation of variability through admixture and introgression. In the face of ongoing global changes and the predicted population increases of the next decades, this plasticity must be ensured and indeed promoted. Understanding its dynamics is crucial in order to repeat and accelerate the process of generating diversity. Building on the most extensive data set of rice genomic variation to date, the 3K RG, we resorted to machine-learning approaches to describe rice genetic diversity in detail. We found evidence of extensive exchanges of genetic material and points of contact between the major groups *Japonica*, *Indica* and *Aus*. Focus on the *circum*Basmati group revealed traces of profuse admixture with contributions of differentiated material in addition to those of the three major groups. The respective components highlighted an ancient origin and a higher affinity with current wild and cultivated Japonica forms from NorthEast India and continental Southeast Asia. The methodology applied presents obvious benefits for the description of genetic variation, for its speed and low computation load. The cartography of genome admixture produced has the potential to become a powerful tool for the understanding and use of rice genetic resources in the future.

P0918: Rice

Reproduction of Rice Transcriptome Dynamics Under Fluctuating Field Environment

Yoichi Hashida, Research Institute for Food and Agriculture, Ryukoku University, Otsu, Shiga, Japan and Atsushi J. Nagano, Center for Ecological Research, Kyoto University, Otsu, Shiga, Japan

For optimizing crop productivity in agricultural field, it is essential to understand the plant response to fluctuating field environment. Numerous studies conducted in controlled laboratory condition have contributed to our understanding of how plants respond to environmental stimuli such as light and temperature. However, plant response to the environmental stimuli in the field are sometimes found to be different from that in controlled condition. To investigate the plant response to the fluctuating field condition in a laboratory, we developed SmartGC, a high-performance growth chamber that can control light, temperature and humidity by 1-minute resolution. By using SmartGC, we aimed to reproduce the diurnal transcriptome dynamics of rice leaf blade in the fluctuating field environment.

Rice seedlings were grown in a growth chamber and then transferred to 3 conditions: field, a growth chamber where light, temperature and humidity were controlled in day/night cycle, and SmartGC where light, temperature and humidity were set to condition simulated from the meteorological data measured in the field experiments. Rice leaf blade was sampled every 2 hours throughout a day and the diurnal transcriptome was determined by RNA-Sequencing analysis.

We successfully simulated field conditions by SmartGC and obtained diurnal transcriptome of rice leaf blade in the 3 conditions. From the results, we will discuss the similarities and differences of transcriptome dynamics in field and simulated condition, and utility of SmartGC for understanding plant response to the fluctuating field environment.

P0919: Rice

Genetically Stable Expression of Brazzein, a New Type of Alternative Sweetener in Transgenic Rice

Yu Jin Jung[1], Akter Sahina[2] and Kwon-Kyoo Kang[1], (1)Hankyong National University, Ansung, South Korea, (2)Department of Horticultural Life Science, Hankyong National University, South Korea

Brazzein is the smallest sweet protein which was isolated from the fruit pulp of *Pentadiplandra* brazzeana Baillon plant, native to tropical Africa. From ancient time, the indigenous people used this plant fruit in their diet to alter the taste of their daily food. It is determined that brazzein is 500 to 2000 times sweeter than sucrose on weight basis and 9500 times sweeter on the molar basis. This unique property has led to increasing interest in this protein. But it is expensive and difficult to produce brazzein other than it's native growing condition which limiting its availability for food products. In this study, we report the high yield production of sweet protein, brazzein in transgenic rice plants. A brazzein driven by the 2 x CaMV 35S promoter was introduced into rice (*Oryza sativa* Japonica) via *Agrobacterium*-mediated transformation. After transformation, 17 regenerated plant lines were obtained and the transgenic lines were confirmed by the PCR amplification. Also, the selected plants lines were gone through Taqman PCR and result showed that 6 T_0 lines were found single copy out of 17 transgenic plants (T_0). Moreover, high and genetically stable expression of brazzein was confirmed by Western blot analysis. These results demonstrate that recombinant brazzein was correctly expressed in transgenic rice plants, and developed a new rice variety with natural sweetener.

P0920: Rice

Biological and Physiological Implications of *qDTY$_{12.1}$* introgression to Basal Growth and Developmental Processes in Rice *Cv. IR64* As Inferred from Transcriptome Signatures

Pushpinder Pal Kaur[1], Naga Bhushana Rao Karampudi[1], Ai Kitazumi[1], Isaiah Pabuayon[1], Nitika Sandhu[2], Margaret Catolos[2], Arvind Kumar[2] and Benildo G. de los Reyes[1], (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)International Rice Research Institute, Los Banos, Philippines

The qDTY$_{12.1}$ is a QTL from the rice cultivar Way Rarem with major effects on yield maintenance under drought stress when introgressed into specific genetic backgrounds. Multi-year trials at IRRI have shown that the optimal effects of qDTY$_{12.1}$ maybe be dependent on or influenced by other minor QTL or network of genes that are brought in the same genetic background through non-targeted introgression. These interactors, which may act either as positive synergists or antagonists, can either cause an enhancing or dragging effects to the QTL. In this initial study, we used the basic comparative panel comprised of the qDTY$_{12.1}$ donor Way Rarem, recipient IR64, and two of their sibling introgression lines for qDTY$_{12.1}$ with contrasting yield under drought to examine transcriptomic differences that may provide better insights on the nature synergistic or antagonistic interactions involved. Transcriptomic profiles for all four lines at vegetative, booting, and grain filling stages under control and drought conditions were compared, first to examine differential expression of homologous genes contained within the qDTY$_{12.1}$ boundary across donor and recurrent parents, and their yield-contrasting introgression lines in IR64. Second, the transcriptome data was interrogated for global gene expression patterns that may provide insights on the network of genes in the genetic background that interact with the qDTY$_{12.1}$ genes in both positive or negative manner to create either an enhancing or dragging physiological effects in low-yielding and high-yielding introgression lines. Results are being used to formulate hypothesis to examine other types of qDTY interactions in different genetic backgrounds.

P0921: Rice

Historical Meiotic Crossover Hotspots Are Associated with DNA Transposons and Chromatin Modification in Rice

Alexandre Marand and Jiming Jiang, University of Wisconsin-Madison, Madison, WI

Recombination plays an integral role in the creation of nascent haplotypes and provides the foundation for genetic variation in sexually reproducing species. Despite this importance, both the extent and determinants of recombination rates across the rice genome have been poorly studied. Here we present the construction of two *Oryza sativa* historical recombination rate maps derived from coalescent estimates using 150 rice genomes. The fine-scale recombination rate maps were validated by comparison with a consensus genetic map composed of three independent synthetic populations and were subsequently used for recombination hotspot identification. Recombination rates varied dramatically on a global scale, with higher rates in the euchromatic arms as defined by gene, retrotransposon and DNA methylation densities, consistent with reports in other plant genomes. Rice recombination hotspots were enriched upstream of gene TSSs and depleted within gene bodies. Coincident with their location upstream of genes, hotspots carried signatures of open chromatin, marked by increased DNase I sensitivity and flanked by activating histone modifications. Discriminative motif discovery in rice recombination hotspots were highly similar to known recombination motifs and reflect sequence content known to exclude nucleosome occupancy. Collectively, these results demonstrate that historical recombination hotspots are determined by fine-scale genomic and epigenomic signatures indicative of open chromatin and *cis* regulatory elements in rice.

P0922: Rice

Tracking the Genome-Wide Outcomes of a Transposable Element Burst over Decades of Amplification

Lu Lu[1], **Jinfeng Chen**[1], Sofia Robb[1], Yutaka Okumoto[2], Jason Stajich[1] and Susan Wessler[1], (1)University of California, Riverside, CA, (2)Graduate School of Agriculture, Kyoto University, Kyoto, Japan

To understand the success strategies of transposable elements (TEs) that attain high copy numbers, we analyzed two pairs of rice (*Oryza sativa*) strains: EG4/HEG4 and A119/A123, undergoing decades of rapid amplification (bursts) of the class 2 autonomous *Ping* element and the nonautonomous miniature inverted repeat transposable element (MITE) *mPing*. Comparative analyses of whole genome sequences of the two strain pairs validated that each pair has been maintained for decades as inbreds since divergence from their respective last common ancestor. Strains EG4 and HEG4 differ by fewer than 160 SNPs and a total of 264 new *mPing* insertions. Similarly, strains A119 and A123 exhibited about half as many SNPs (277) as new *mPing* insertions (518). Examination of all other potentially active TEs in these genomes revealed only a single new insertion out of ~40,000 loci surveyed. The virtual absence of any new TE insertions in these strains outside of the *mPing* bursts demonstrates that the *Ping/mPing* family gradually attains high copy numbers by maintaining activity and evading host detection for dozens of generations. Evasion is possible because host recognition of *mPing* sequences appears to have no impact on initiation or maintenance of the burst. *Ping* is actively transcribed and both *Ping* and *mPing* can transpose despite methylation of terminal sequences. This finding suggests that an important feature of MITE success is that host recognition does not lead to the silencing of the source of transposase.

P0923: Rice

The Recognition Assay of Termination Codon By *ESP1/ Eukaryotic Releases Factor-1 (eRF1)* in Rice

Ammar Elakhdar[1], Tomokazu Ushijima[1], Masako Fukuda[1], Noriko Yamashiro[1], Yasushi Kawagoe[2] and Toshihiro Kumamaru[1], (1)Institute of Genetic Resources, Faculty of Agriculture, Kyushu University, Fukuoka, Japan, (2)Division of Plant Sciences, National Institute of Agrobiological Sciences, Tsukuba, Japan

Rice prolamine polypeptides are translocated into the ER-lumen where they indigestible proteins that decrease nutritional value of rice protein. We identified the *esp1*-mutant which accumulate low amount of 13KDa (*CysP*) prolamin molecules with the processed into neutral to acidic subunits. The *ESP1 gene* (Os07g0587400) with base substitution to *eRF1* is exhibit UGA-stop codon, defined to *ESP1-eRF1*. This study aims to investigate the functional of *ESP1-eRF1* for accumulation of *CysP* prolamin in the rice endosperm. We defined the *CysP* prolamin genes that controlled by *ESP1-eRF1* where located on chromosome-5, preferred UAA stops and characterized as free-cysteine residues. We hypothesize that the *ESP1-eRF1* might be selectively recognized UAA stop codons. Transgenic luciferase constructs with UAA and UAG were decreased the mRNA fold-change in *esp1*-mutant compared with their wild-type. However, there is no *CysP* prolamin gene with UAG stops. The significant decrease is due to the loss of function, suggesting that *ESP1-eRF1* along with other three *eRF1* paralogs in rice genome depending on the molecular species is involved in reducing the accumulation of *CysP* prolamin. The microarray analysis of six-growth stages illustrated that, out of 630 *ESP1-eRF1* regulated genes coding for knowing protein; 19, 33 and 48% of genes with UAA, UAG and UGA stops, respectively. Five *CysP* prolamin genes exhibited double stop codons (UAA-UGA) on the mRNA, with amino acids composition Gly-Val-Leu in C-terminal position. We discussed the efficiency of *ESP1-eRF1* translation termination that influenced by local contexts surrounding stop codons act synergistically to involved the accumulation efficiency of *CysP* prolamin.

P0924: Rice

Distinct Gene Expression Signatures Leading to Transgressive Phenotype in *Oryza sativa* L. Shows Novel Avenues for Crop Improvement

Isaiah Pabuayon[1], Ai Kitazumi[1], Rakesh Kumar Singh[2,3], Glenn B. Gregorio[2] and Benildo G. de los Reyes[1], (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)International Rice Research Institute, Los Baños, Laguna, Philippines, (3)International Rice Research Institute, Los Banos, Philippines

Transgressive segregation is a phenomenon creating individuals displaying enhanced phenotypes relative to the parents. We hypothesized that through repeated recombination, network reconfiguration could occur with multiple layers of control, resulting in optimal gene combinations, novel physiological properties, and stronger phenotypes. This is especially beneficial in creating lines for abiotic stress tolerance, as a multifaceted response is essential for adapting and responding to stress. Particularly, reduction of salt toxicity is often addressed by looking for means of Na+ exclusion through the activities of cellular pumps. In this study, an extensive phenotypic evaluation for salt tolerance was conducted across recombinant inbred line population (RIL-F_8) from a cross between IR29 and Pokkali. Results showed wide differences in multiple aspects of stress physiological responses across the population. One of the selected transgressive lines only showed middling performance in terms of Na^+ ion exclusion, and this led us to hypothesize that this mechanism is not the sole reason for its superior stress tolerance properties. Transcriptome analysis revealed novel transcriptomic signatures in the said salt-tolerant transgressive line. Here, we describe the potential roles of a novel regulatory network that is uniquely configured in a transgressive salt-tolerant progeny between IR29 and Pokkali, which appears to configure an osmotic adjustment potential that is above and beyond what can be achieved by the known Na+ exclusion mechanisms in other non-transgressive but salt-tolerant progenies. We also describe the identification of the likely hubs for the reconfigured networks, which may rely not only on sequence variation, but also through epigenetic means.

P0925: Rice

A Functional Study on Pyramiding *Sub1* and *Pup1*: Compatibility Issue of Two QTLs

Joong Hyoun Chin[1], Na-Hyun Shin[2], Jae-Hyuk Han[2], Hee-Jong Koh[3] and Soo-Cheul Yoo[2], (1)Graduate School of Integrated Bioindustry/Sejong University, Seoul, South Korea, (2)Hankyong National University, Ansung, South Korea, (3)Seoul National University, Seoul, Korea, Republic of (South)

More than 40% of rice cultivation area are subject to abiotic stresses. Pyramiding useful genes/QTLs into elite variety is a promising strategy to develop tolerant varieties against multiple abiotic stresses. However, some genes/QTLs may not functionally compatible when they are introgressed into the same variety. In this study, as one example, we have tested the functional compatibility of *Sub1* and *Pup1*, major QTLs for tolerance to submergence and phosphate (P)-deficiency conditions, respectively. Phenotypic analysis revealed that IR64-*Sub1*-*Pup1* (SP1) plants harboring both *Sub1* and *Pup1* QTLs showed significant tolerance to submerged conditions, similarly in IR64-*Sub1*, while SP1 plants failed to tolerate to P-deficiency conditions; only IR64-*Pup1* showed P-deficiency tolerance. In submergence condition, the expression levels of *Sub1A* (*Sub1*) and *OsPSTOL1* (*Pup1*) were not significantly different among IR64-Sub1, IR64-Pup1 and SP1 plants. On the other hand, Under P- condition, the expression level of *OsPSTOL1* was not different from that under P+ condition. However, the expression of S*ub1A* in SP1 was highly increased under P- condition although SP1 didn't show tolerance. These results suggest *Pup1* compromises the *Sub1* function in submerged conditions while *Sub1* suppresses the function of *Pup1* in P-deficiency condition, possibly by working with negative effect overwhelming positive function of *Pup1*. In conclusion, *Sub1* and *Pup1* are functionally compatible in terms of submergence tolerance but not in P-deficiency conditions. The success of combining useful traits would depend on the phenotypic co-expression of QTLs in the target environments. Breeding purpose should be carefully setup with information of QTL/gene compatibility.

P0926: Rice

Characterization of a New Gene Controlling Leaf Senescence Using Progeny from an Interspecific Cross in Rice

Kyu-Chan Shim[1], Luong Ha[2], Sun-Ha Kim[2], Yun-A Jeon[1], Hyun-Sook Lee[1] and **Sang-Nag Ahn**[3], (1)Dept. of Agronomy, Chungnam National University, Daejeon, Korea, Republic of (South), (2)Chungnam National University, Daejeon, South Korea, (3)Dept. of Agronomy, Chungnam National University, Daejeon, South Korea

Analysis of quantitative trait loci (QTL) controlling chlorophyll content was conducted using introgression lines (CR7501 and CR2002) derived from a cross between the *Oryza sativa japonica* cultivar 'Hwaseong' as a recurrent parent and wild species *O. grandiglumis* as a donor

parent. These two lines showed higher chlorophyll content than Hwaseong. For QTL analysis, we costructed 58 F3 and 38 F4 lines from the cross CR7501/Hwaseong. SSR markers were used for genotyping the lines. One-way ANOVA indicated the presence of a QTL for chlorophyll content (*qCC2*) on chromosome 2 and *qCC2* explained 24.6% of the phenotypic variance. To examine whether *qCC2* is also involved in senescence, a series of dark-induced senescence (DIS) experiments were conducted. Detached leaves from Hwaseong and CR2002 were incubated in 3mM MES buffer (pH 5.8) at 27 °C under complete dark condition. CR2002 showed higher chlorophyll content with delayed senescence than Hwaseong. To know whether *qCC2* maintains leaf functionality during DIS, ion leakage test and Fv/Fm measurement were performed. The Fv/Fm value displayed significant difference between CR2002 and Hwseong at 6 days after incubation, while ion leakage rate was not significantly different. These results might imply that *qCC2* is associated with chlorophyll content and stay-green phenotype. The *qCC2* QTL region harbors *GW2* locus encoding the E3 ubiquitin ligase controlling grain size. To know the possible relationship between the *GW2* activity and senescence, *gw2*-knockout mutant (*gw2*-KO) and the wild type plants were incubated under dark condition. *gw2*-KO showed delayed senescence in DIS, suggesting that *GW2* is possibly related to stay-green phenotype.

P0927: Rice

Uncovering Distinctive Spatio-Temporal Transcriptome Signatures across Genotypes, Stress Regimes, and Developmental Stages in Rice: Assumptions and Strategies

Naga Bhushana Rao Karampudi, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

Understanding spatio-temporal changes in gene expression at a global scale especially across a population of recombinants that share common parents, and across treatment regimes, and developmental stages could be prone to lots of noise, non-biologically meaningful variation, and other subtleties that introduce different biases and errors in interpretation. Here we describe the overall strategy that we employed in an effort to reveal meaningful similarities and differences in terms of transcriptome signatures across recombinant inbreds and introgression lines of rice and their parents in context of novel phenotypes. Using two RNA-Seq datasets, one from a F8 recombinant inbred line population and another from a near isogenic line population under distinct experimental conditions, we conducted a parallel comparison of transcriptome signatures using two novel computational approaches. The Percentage Difference method approximates the change in transcript abundance and relies on an optimized cut-off based on robust statistics. The Hybrid method uses quartile on top of the percentage difference. The Percentage Difference method provides better resolution for analyzing datasets within a given developmental stage, whereas the Hybrid method provided better resolution for comparing across developmental stages. The quartile assumption in hybrid method takes care of inherent transcript abundance variation or fluctuation associated with growth or developmental stages. Both methods are very efficient in terms of removing the noise and extracting meaningful biological conclusions. We recommend Hybrid method because of its wide applicability regardless of experimental conditions and efficiency in identifying the transcriptome signatures.

P0928: Rice

Application of High-Throughput SNP Genotyping Method to QTL for Allelopathy Traits in Rice

Jisu Choi , Mijeong Kim, Kyoungjin Lee, Nayoung Son, Joohyun Lee and Ill-Min Chung, Konkuk University, Seoul, South Korea

Weeds are major damaging factors in rice production and quality because of competition between weeds and crops on light, nutrients, water and space. The Allelopathy to suppress the surrounding weeds by exporting chemicals called allelochemical may be additional rice weed management method. Introducing the allelopathy trait into the parents line for F_1 seed production in rice will be beneficial.

In this study, QTL analysis of allelopathy was conducted. A total of 160 of F_8 RILs developed from the cross between Nongan(low allelopathic cultivar) and Sathi(high allelopathic cultivar) were used . the performance of allelopathy were evaluated using 'ECAM(Equal Compartment Agar Method)', where the root length, shoot length, root weight and shoot weight of lettuce cultivated with the RILs were measured. In order to identify the location of QTLs related to allelopathy, genotyping was carried out using Fluidigm chips with 120 SNP markers which distinguish *japonica* and *indica* type.

The distribution of the performance was followed as normal distribution and the result of QTL analysis, candidate QTLs in chromosome 3 and chromosome 6 were detected. These QTLs are related to Root Length and the range of LOD score was from 2.9 to 5.9 and PVE(phenotypic variance explained) range was from 2.2% to 11.6%.

Identifying genes related allelopathy using RIL populations requires much time and labor from phenotyping to genotyping. However, ECAM(Equal-compartment-agar-method) which is laboratory test and high-throughput SNP genotyping method using Fluidigm chips can analyze candidate genes quickly and accurately.

P0929: Rice

Duplication of an Upstream Silencer of *FZP* Increases Grain Yield in Rice

Xufeng Bai[1], Yong Huang[2], Yong Hu[2], Haiyang Liu[2], Bo Zhang[2], Cezary Smaczniak[3], Gang Hu[2], Zhongmin Han[2] and YongZhong Xing[1,4], (1)Huazhong Agricultural University, Wuhan, China, (2)Huazhong Agricultural University, China, (3)Humboldt-Universität zu Berlin, Germany, (4)National Key Laboratory of Crop Genetic Improvement, Wuhan, China

Transcriptional silencer and copy number variants (CNVs) are associated with gene expression. However, their roles in generating phenotypes have not been well studied. In this study, we identified a rice quantitative trait locus (QTL), *SGDP7* (*Small Grain and Dense Panicle 7*). *SGDP7* is identical to *FZP* (*FRIZZLE PANICLE*), which represses the formation of axillary meristems. An 18-bp fragment, named CNV-18bp, was inserted ~5.3 kb upstream of *FZP*, thus resulting in a tandem duplication in Chuan 7. The CNV-18bp duplication repressed *FZP* expression, prolonged the panicle branching period and increased grain yield by more than 15% by substantially increasing the number of spikelets per panicle (SPP) and slightly decreasing 1000-grain weight (TGW). The transcription repressor OsBZR1 binds the CGTG motifs in CNV-18bp and represses *FZP* expression, indicating that CNV-18bp is the upstream silencer of *FZP*. These findings showed that CNVs of the silencer coordinate a trade-off between SPP and TGW by fine-tuning *FZP* expression, and balancing the trade-off would enhance yield potential.

P0930: Rice

Genome-Wide Association Study (GWAS) of Coleoptile and Mesocotyl Elongation in Rice (Oryza sativa L.)

Sandeep A Sakhale, University of Illinois, Department of Crop Sciences, Urbana, IL, Georgia C. Eizenga, USDA-ARS Dale Bumpers National Rice Research Center, Stuttgart, AR, Alexander E. Lipka, University of Illinois, Urbana, IL, Arvind Kumar, International Rice Research Institute, Los Banos, Philippines and Erik J. Sacks, Dept of Crop Sciences, University of Illinois, Urbana, IL

Direct-seeding of rice without prior pre-germination is gaining popularity in rice growing countries because it needs less water and labor than transplanting rice seedlings. Slow emergence and poor seedling establishment of direct-seeded rice are the primary drawback of this method. Potential shoot length is important for successful direct seeding of rice, especially if seeds are sown deeply to ensure access to moisture. To assess variation among cultivars and identify genomic regions associated shoot length, we studied 238 *Japonica* accessions from Rice Diversity Panel (RDP1). Two complementary experiments were conducted, with one in an incubator (Expt 1) and the other in a greenhouse (Expt 2). For Expt 1, seed was placed on a moist paper towel (ragdoll method) rolled up and placed in a tube for 5 days in the dark at30 °C. Subsequently, seedlings were photographed and coleoptile length was measured with ImageJ softwareFor Expt 2, seeds were planted 8 cm deep in 3.8 x 18.4 cm pots (Cone-tainers) filled with a 1:1:1 mixture of field soil, peat moss, and sand, and cultured for 21 days at 29/25 °C day/night. Highly significant differences among the accessions were observed for coleoptile length in Expt 1 (0.7-5.2 cm range), and total shoot length (4.2-49 cm range) and days to emergence (4-21 days) in Expt 2. Genome wide association analysis was conducted using HDRA dataset consisting 700,000 SNPs. Present study will enrich our knowledge about the genomic regions associated with the coleoptile and mesocotyl traits, which will help us to develop new cultivars with improved adaptation to direct seeding.

P0931: Rice

Systems Genetics Identifies a Novel Regulatory Domain of Amylose Synthesis

Roslen S. Anacleto[1], Vito M. Butardo Jr.[2], Sabiha Parween[1], Irene Samson[1], Krishna de Guzman[1], Crisline Mae Alhambra[1], Gopal Misra[1] and Nese Sreenivasulu[1], (1)International Rice Research Institute, Los Baños, Philippines, (2)Swinburne University of Technology, Victoria, Australia

A deeper understanding of the regulation of starch biosynthesis in rice (*Oryza sativa* L.) endosperm is crucial in tailoring digestibility without sacrificing grain quality. In this study, significant association peaks on chromosomes 6 and 7 were identified through a genome-wide association study (GWAS) of debranched starch structure from grains of a 320 indica rice diversity panel using genotyping data from the high-density rice array. A systems genetics approach that interrelates starch structure data from GWAS to functional pathways from a gene regulatory network identified known genes with high correlation to the proportion of amylose and amylopectin. A SNP in the promoter region of *Granule Bound Starch Synthase I* was identified along with seven other SNPs to form haplotypes that discriminate samples into different phenotypic ranges of amylose. A GWAS peak on chromosome 7 between LOC_Os07g11020 and LOC_Os07g11520 indexed by a nonsynonymous SNP mutation on exon 5 of a bHLH transcription factor was found to elevate the proportion of amylose at the expense of reduced short-chain amylopectin. Linking starch structure with starch digestibility by determining the kinetics of cooked grain amylolysis of selected haplotypes revealed strong association of starch structure with estimated digestibility kinetics. Combining all results from grain quality genomics, systems genetics, and digestibility phenotyping, we propose target haplotypes for fine-tuning starch structure in rice through marker-assisted breeding that can be used to alter the digestibility of rice grain, thus offering rice consumers a new diet-based intervention to mitigate the impact of nutrition-related noncommunicable diseases.

P0932: Rice

Whole Genome Sequencing-Based Association Study to Unravel Genetic Architecture of Cooked Grain Width and Length Traits in Rice

Gopal Misra[1], Saurabh Badoni[1], Roslen S. Anacleto[1], Andreas Graner[2], Nickolai Alexandrov[1] and Nese Sreenivasulu[1], (1)International Rice Research Institute, Los Baños, Philippines, (2)Leibniz Institute of Plant Genetica and Crop Plant Reserch, Gatersleben, Germany

In this study, we used 2.9 million single nucleotide polymorphisms (SNP) and 393,429 indels derived from whole genome sequences of 591 rice landraces to determine the genetic basis of cooked and raw grain length, width and shape using genome-wide association study (GWAS). We identified a unique fine-mapped genetic region GWi7.1 significantly associated with cooked and raw grain width. Additionally, GWi7.2 that harbors GL7/GW7 a cloned gene for grain dimension was found. Novel regions in chromosomes 10 and 11 were also found to be associated with cooked grain shape and raw grain width, respectively. The indel-based GWAS identified fine-mapped genetic regions GL3.1 and GWi5.1 that matched synteny breakpoints between indica and japonica. GL3.1 was positioned a few kilobases away from GS3, a cloned gene for cooked and raw grain lengths in indica. GWi5.1 found to be significantly associated with cooked and raw grain width. It anchors upstream of cloned gene GW5, which varied between indica and japonica accessions. GWi11.1 is present inside the 3′-UTR of a functional gene in indica that corresponds to a syntenic break in chromosome 11 of japonica. Our results identified novel allelic structural variants and haplotypes confirmed using single locus and multilocus SNP and indel-based GWAS.

P0933: Rice

Regulation of Grain Size/Shape and Grain Chalkiness in Rice

Yibo Li, National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan, China

Rice grain size or shape, having a very high heritability, is not only one of the most important target traits of grain yield, appearance quality, domestication and breeding, but also an ideal model of morphogenesis and development of organ shape in higher plants; How horizontal and vertical growth of grain co-regulate the final grain shape is a basic science question. To focus on this question, 475 enhancers or suppressors of grain shape with 8 directions and 328 mutants of various grain-chalkiness types were identified, and the genes and networks are being isolated employing the MutMap method. Moreover, 72 rare natural variations (enhancers or suppressors) of grain shape with 8 directions were

discovered, and 8 novel QTLs for each direction of grain shape were identified by BSA-based RICE6K arrays, confirmed by screening the NILs and cloned by both map-based cloning strategies and candidate gene methods. Regulations of grain shape by two major genes *GS3* and *GW5* is being done. And then grain-shape modelling in rice will be conducted in the future.
Grain chalkiness in rice, a chalky texture of endosperm, is a highly undesirable quality trait of the grain for human food that greatly affects grain appearance, head rice yield, marketing values and cooking and eating qualities of grain. So it represents a major problem in many rice-producing areas of the world, but its regulation mechanism is unknown. To focus on this question, large-scale construction of ILs for various gain chalkiness types based on phenotypes of parents and genetic populations were conducted, including 55 *indica* / *indica* rice populations, 56 *indica* / *japonica* rice, 52 *japonica* / *japonica* rice populations, and 38 high-temperature tolerance / sensitive rice, each population with a single-locus segregating pattern. Moreover, we performed GWAS for all grain chalkiness traits with four-year field replications, and detected 29 stable QTLs, and 15 major QTLs sensitive to temperature using the difference values of the two responsive chalky traits between the high (33°C/27°C, 2013 at Wuhan) and low temperature (27°C/21°C, 2014 at Wuhan) natural environments. 9 novel and major QTLs for grain chalkiness were identified by BSA-based RICE6K arrays consistent with GWAS results, confirmed by screening the NILs, and have being isolated by both map-based cloning strategies and candidate gene methods, to uncover the general regulating mechanisms for the formation of grain chalkiness and quality in rice.
Key words: Rice, Grain shape, Grain chalkiness, Yield, quality, Regulation

P0934: Rice
Genome-Wide Characterization and Expression Analysis of the Germin-like Protein Family in Rice and Arabidopsis
Lu Li, Nanjing Agricultural University, China, Nanjing, China
Previous studies have shown that germin-like proteins (GLPs) are present ubiquitously in rice and Arabidopsis. However, the understanding regarding their role in development and abiotic/biotic stress resistance remains limited. In the present study, we report genome-wide identification, characterisation, subcellular localization, enzyme activity, and expression analysis of the GLP gene family in rice and Arabidopsis to study their functions. In total, 43 and 32 GLPs in the rice and Arabidopsis genome were identified based on a systematic analysis, respectively. The GLP genes were clustered into six clades based on phylogenetic analysis, and many stress and developmental-related cis-elements were detected in promoters of GLP genes. In addition, subcellular location and superoxide dismutase (SOD) analysis demonstrated that the random selected OsGLP genes on chromosomes 8 and 4 of rice were expressed in the cell wall with SOD activity. Overall, our results showed that tandem duplication events, especially the clusters of tandem duplication genes on chromosome 8 in rice, play a major role in expansion of the GLP family and thus increase our understanding of the role of the GLP family in abiotic/biotic stress and development.

P0935: Rice
Effect of 60Co~γ Ray Radiation on Changes of Chromatin States in Rice
Xiucai Pan, Yuan Fang and **Wenli Zhang**, State Key Laboratory for Crop Genetics and Germplasm Enhancement, Nanjing Agriculture University,, Nanjing, China
Involvement of Histone modifications are in the DNA damage response has been reported in non-plant systems, including mammals and yeast. However, how chromatin dynamics, either an individual mark or combined chromatin states, participate in regulating differentially expressed genes in the plant DNA damage response is still understudied. We conducted RNA-seq and ChIP-seq to demonstrate that differentially expressed genes (DEGs), in response to ionizing radiation (IR), might be involved in different pathways responsible for the DNA damage response. Moreover, chromatin structures associated with promoters, exons and intergenic regions are significantly affected by IR. Most importantly, either an individual mark or a certain chromatin state was found to be highly correlated with the expression of up-regulated genes. In contrast, only the chromatin states, as opposed to any individual marks tested, might be responsible for regulating expression of the down-regulated genes. Our findings indicate that distinct epigenetic mechanisms modulate the transcription of IR-related DEGs. Either chromatin states or distinct histone dynamics may act sequentially or in combination in regulating up-regulated genes, but the complex chromatin structure is mainly responsible for the expression of down-regulated genes. Thus, our study provides new insights into how up- and down-regulated genes are epigenetically regulated at the chromatin levels, thereby helping us to understand distinct epigenetic mechanisms that function in the plant DNA damage response.

P0936: Rice
Characterizing a Rice Diversity Panel with a 7K SNP Chip and Flowering Time Evaluation
Karina Y. Morales[1], Stephon Warren[1], John Carlos I. Ignacio[2], Yuxin Shi[3], Rodante Tabien[4], Tobias Kretzschmar[5], Susan McCouch[3] and Michael J. Thomson[1], (1)Texas A&M University, College Station, TX, (2)International Rice Research Institute, Los Baños, Philippines, (3)Cornell University, Ithaca, NY, (4)Texas A&M AgriLife Research, Beaumont, TX, (5)International Rice Research Institute, Los Banos, Philippines
Rice (Oryza sativa L.) is an essential food crop with demands for increased yield as it provides the daily caloric intake of over 50% of the world's growing population. Flowering is one of the most sensitive stages of rice growth and is highly variable among varieties and across environments. In Texas, farmers often desire early flowering varieties as these can avoid peak temperatures of the summer months and give sufficient time for the ratoon crop to mature before the cold temperatures of winter begin. This experiment took place at the Texas A&M AgriLife Research Center in Beaumont, TX where 208 rice varieties of diverse origins were planted in spring 2017 and were grown through the summer of 2017. Beginning approximately 50 days after planting, notes were collected once a week on flowering percentage to estimate days to 50% flowering. Each variety was genotyped using the Illumina 7K rice SNP chip developed at Cornell University. This project aims to identify genetic loci which contribute to extremely early and late flowering time. Upon identifying these loci, we will use the CRISPR/Cas9 genome editing system to validate candidate genes in diverse genetic backgrounds to gain a better understanding of how each locus may contribute to days to heading in rice. Ultimately, the improved knowledge on manipulating flowering time genes will lead to more precise tools to provide early flowering in any genetic background for each target environment.

P0937: Rice

Rice *wrky-mediated stay-green1* Regulates Chlorophyll b Degradation during Senescence

Na-Hyun Shin, Jae-Hyuk Han and **Soo-Cheul Yoo**, Hankyong National University, Ansung, South Korea

Chlorophyll breakdown is an vital catabolic process of leaf senescence and fruit ripening as it allows for recycling of nitrogen and other nutrients. Through the screening of T-DNA insertional lines, we found a stay-green mutant, *wrky-mediated stay-green1* (*wms1*), showing delayed leaf senescence during both natural and dark-induced leaf senescence. Although chlorophyll degradation was remarkably inhibited, photosynthetic capacity diminished more or less normally like wild-type, indicating that *wms1* is a non-functional stay-green mutant. By ultrastructural analysis of *wms1* chloroplasts, very thick and wide grana stacks of thylakoid membranes were observed in senescent *wms1* mutant. In addition, chlorophyll b was highly accumulated as well as Light Harvesting Chlorophyll b (Lhcb) proteins remained abundantly in the senescent leaves of *wms1*, similar to other non-functional stay-green mutants, *sgr* and *nyc1*. We further tested whether *wms1* stays green under several hormone treatments and found that *wms1* still displayed strong greenness under MeJA and ACC (ethylene), whereas partial leaf yellowing was observed under ABA, indicating that *WMS1* plays a important role in the signaling of MeJA and ethylene-induced senescence but not in ABA. Further analysis is necessary to identify a target of WMS1 to elucidate such a strong regulation mechanism of transcription factor-mediated senescence.

P0938: Rice

Genome Wide Investigation and Expression Analysis of AP2 Transcription Factor Sub-Family Reveals Its Evolution, Expansion and Regulatory Role in Developmental Processes in indica Rice

Syed Adeel Zafar, Institute of crop science, Chinese academy of agricultural sciences, Beijing China, Haidian, China and Babar Ijaz, China Agricultural University, Beijing, China., Beijing, China

A comprehensive genome wide investigation was carried out which enabled us to identify 26 non-redundant genes encoding AP2 transcription factor subfamily in Indica rice. These genes were un-evenly distributed among eleven out of twelve rice chromosomes. Phylogenetic analysis of OsAP2 TF family members grouped them into three major classes clearly indicating the paralogous genes. Segmental duplication seemed to be principal route of evolution of rice AP2 genes, supporting the higher positive selection pressure during evolution. The inference of duplication dates of paralogous gene pairs suggested that segmental duplications originated about 2.36 to 101.40 million years ago (Mya); while tandem duplication about 13.66 Mya in AP2 subfamily of O. sativa. Conserved domain analysis revealed that most of the conserved domains, apart from the AP2 domain, are exclusively distributed among the specific clades in the phylogenetic tree. MicroRNA study indicated that osa-miR172 family was potentially targeting four AP2 members in rice. Expression analysis illustrated that different OsAP2 genes have tissue specific expression. However, the genes show high expression pattern in seed and endosperm tissues, suggesting their putative role in seed development. Based on orthologous genes from Arabidopsis, several AP2 genes in rice were identified to be involved in different growth and development related functions like floral growth and patterning, organ size control, biosynthesis of storage compounds during seed development, ABA signaling and seedling germination. This study will serve as a foundation for further molecular characterization of OsAP2 genes in development and stress response.

P0939: Rice

QTLs Related to Stigma Exertion Trait of Rice Derived from *Oryza rufipogon* 'W0120'

Citra Bakti, University of Tsukuba, Tsukuba, Japan and Junichi Tanaka, The Institute of Crops Science, NARO (NICS), Tsukuba, Japan

Many breeding programs of autogamous crop species become slow down and stagnant compared to the allogamous crops such as maize. A novel breeding methodology based on efficient out-crossing and transgenic male sterility utilization was proposed by Tanaka (2010) to overcome the problem. In the outcross-based breeding system, improvement of plant productivity by outcrossing is desirable, and stigma exertion trait can play a key role to increase the out-crossing fertility in rice. This trait is also crucial in the F_1 production of hybrid cultivars, and there are many studies were proceeding. We focused on *Oryza rufipogon*, an ancestor species of rice, because it can be expected to keep the wild-type allele of many domesticated genes. We screened the material 'W0120' based on the photo images in *OryzaBase* database, as an accession with typical and clear stigma exertion trait. We performed a quantitative traits loci (QTL) analysis for stigma exertion trait using F_2 population derived from crossing between temperate japonica cultivar 'Akidawara' and 'W0120'. The F_2 population had a wide range for stigma exertion rate, range between 0% and 90%. We found the minor QTLs on chromosome 2 and 11; and the major QTLs on chromosome 3 and 8, respectively. In the next step, we are going to clone the responsible gene of the QTLs to facilitate the outcrossing for the novel breeding methodology using male sterility.

P0940: Rice

Mining the Genetic Variation for Yield Component Traits from a Collection of *O. rufipogon* Accessions and its Introgression into Cultivated Rice

Palvi Malik, Ohio Agricultural Research and Development Center, Ohio State University, Wooster, OH; Punjab Agricultural University, Ludhiana, India and Dharminder Bhatia1, Kumari Neelam1, Bharat Yadav1 and Clay H. Sneller2 1School of Agricultural Biotechnology (PAU, India); 2 Ohio Agricultural Research and Development Center (OSU)

The potential of wild germplasm being a source of useful genetic variability has been well documented. Therefore, a set of 370 different accessions of *Oryza rufipogon* (wild progenitor of cultivated rice, *Oryza sativa*) from different geographical origins was evaluated. All the accessions were phenotyped in replications for two years for yield component traits namely plant height, culm thickness, panicle length, number of primary branches per panicle, grain length, grain width and thousand grain weight. Genotyping of all the accessions using GBS identified 119,000 SNPs, out of which 59,335 SNPs were retained after filtering for missing data point <10 % and 0.05 MAF. The analysis of population structure by fastStructure program revealed three genetically distinct groups. GWAS studies conducted using FarmCPU revealed significant marker-trait associations for grain length, grain width, panicle length, number of primary branches and culm thickness. In order to

dissect and incorporate these alleles into breeder's pool, wide-cross hybridization has been attempted and a total of 10,632 BC_2F_1 individuals and 128,000 BC_1F_2 individuals belonging to 35 different families have been generated. A subset of the backcross population has been selected based on kinship, subjected to GBS, and phenotyped for productivity related traits. This subset will be used as training population to calculate the marker effects and initiate genomic selection. The lines with best Genomic Estimated Breeding Values will be selected, hybridized among themselves to initiate the next cycle of genomic selection.

P0941: Rice

A Rice Methionine-R-Sulfoxide Reductase, OsMSRB5, Is Required for the Defense Against Excess Copper and Methyl Viologen

Tengwei Xiao, Mengmeng Mi, Changyong Wang, Zhubing Hu, Yan Xia and Zhenguo Shen, College of Life Sciences, Nanjing Agricultural University, China, Nanjing, China

Methionine sulfoxide reductases (MSRs), a family of enzymes catalyzing the conversion of methionine from its oxidized form into reduced form, has a pivotal role against stresses. In the present study, we found the crucial role of rice OsMSRB5 in the adaption to Cu stress. *OsMSRB5* was mainly expressed in leaves, with low transcriptional levels of *OsMSRB5* observed in seeds, stems, and roots, and also was induced by copper (Cu) and methyl viologen (MV) treatment. Subcellular localization analysis indicated that OsMSRB5 was localized to the cytoplasm. The functions of OsMSRB5 were analyzed through ectopic expression in *Escherichia coli* and functional disruption in rice. Ectopic expression of *OsMSRB5* conferred *E. coli* cells the higher resistance to Cu and MV exposure. In rice, the *osmsrb5* mutants showed the decreased Cu and MV tolerance compared with wild-type rice seedlings. Enzymatic activity assay *in vitro* showed that OsMSRB5 had ability to reduce free Met-R-SO and dabsyl-MetSO (a protein-bound-like MetSO) to Met and dabsyl-Met, respectively. Taken together, our results suggest that OsMSRB5 plays an important role in defense against Cu and MV toxicity.

P0942: Rice

GWA Mapping of Oryza sativa Median Lethal Low Temperature (LT50) QTL Using the USDA Rice Diversity Panel 1 (RDP1)

Naoki S. Shimoyama, Marquette University, Milwaukee, WI

Improving crop productivity is vital to meet the needs of a growing world population. Rice is a staple crop feeding approximately half of the world population. Cold stress is a major factor limiting rice productivity due to its tropical origin. We measured seedling survivability at different temperatures for different rice cultivars with varying degrees of cold sensitivities to determine the Median Lethal Low Temperature (LT50) and used it for quantitative trait locus (QTL) mapping to identify potential mechanisms involved in cold stress tolerance. We calculated the LT50 of more than 300 cultivars from the USDA Rice Diversity Panel 1 (RDP1), because its cultivars were genotyped using the High Density Rice Array (HDRA), and 700,000 SNPs are currently available. The RDP1 and HDRA were used in an effort to produce data with great breadth in diversity and resolution, respectively. Using the open source genome-wide association (GWA) mapping pipeline based on the HDRA data, we were able to identify numerous QTL for the LT50 trait. The LT50 phenotypic data will also be used to bin the cultivars with similar median temperatures. By conducting GWA mapping and QTL analysis between these bins, we expect to identify mechanistically significant genes associated with certain LT50 bins. With this analysis, we expect to show that different LT50 bins may employ unique cold tolerance mechanisms and pathways. Genetic pathways linked to membrane composition, lipid synthesis, mechanosensing, chemosensing, and metabolism will be discussed.

P0943: Rice

Global Alternative Polyadenylation Dynamics in Response to Biotic and Abiotic Stresses in Rice

Congting Ye, Qian Zhou and **Qingshun Quinn Li**, Xiamen University, Xiamen, China

Alternative polyadenylation (APA) is a main source of transcriptome diversity, and is extensively involved in the post-transcriptional regulation of gene expression in many biological processes. To take the advantage of the unprecedented underutilized RNA-seq data, we developed a tool called APAtrap - a novel bioinformatics tool developed to identify APA sites and APA dynamics from free RNA-seq data. Using APAtrap, we investigated the roles of APA in the stress-responsive mechanism of rice under abiotic or biotic stimuli. Hundreds of genes with dynamic APA usage were found in 6 experiment datasets of rice under different stress conditions like drought, cadmium, and heat shock abiotic stresses, and bacterial blight, blast, and rice stripe virus biotic stresses. Functional pathway enrichment analysis also uncovered that these dynamic APA events were significantly enriched in many crucial pathways, such as reactive oxygen species scavenging pathways, plant hormone signal transduction and MAPK signaling pathways, indicating the widespread involvement of APA in organism's stress-responsive regulation. Interestingly, differential usages of APA in transcription factors genes such as *RING zinc-finger protein genes* and signaling regulatory genes such as *calmodulin* were relevant with the crosstalk between biotic and abiotic stress-responsive network of rice. Global analysis also indicated that genes with significant differential APA usage tend to be downregulated when mRNA isoforms with longer 3′ UTRs are more abundant. Overall, this study provides a comprehensive overview of APA regulation and complexity in stress-sensitive and stress-tolerant rice cultivars under abiotic and biotic stresses, a previously overlooked phenomenon.

P0944: Rice

Genetic Improvement of Nitrogen Utilization in Rice

Jianru Zuo, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Crops, nitrogen directly determines productivity and biomass, and the improvement of nitrogen utilization efficiency (NUE) is therefore a major challenge in modern agriculture. However, the regulatory mechanism of NUE is
poorly understood. Here, we report the characterization of *are1*. *ARE1* is a highly conserved gene, encoding a chloroplast-localized protein. Loss-of-funciton mutations in *ARE1* cause delayed senescence, the prolonged photosynthetic activity, and the enhanced NUE, thereby increasing grain yield by 10%-20% under nitrogen-limiting conditions.

P0945: Rice

Characterising Agronomic Traits in African Rice Using Sequencing-Based Genotyping

Hayba Badro[1], Marie-Noelle Ndjiondjop[2], Agnelo Furtado[1] and Robert J. Henry[1], (1)University of Queensland/QAAFI, Brisbane, Australia, (2)Africa Rice Center, Cotonou, Benin

In an attempt to preserve many of the essential agronomic traits of both of the cultivated rice species, *Oryza glaberrima* (African rice) and *Oryza sativa* (Asian rice), a promising hybrid of both species has been created. With the advance in high-throughput sequencing platforms, sequencing to identify genes associated with these traits has become an attractive approach to rice breeding particularly for African rice which is a source of important genetic diversity. This study used a BC2F8 population developed by crossing WAB56-104 (*O. sativa*) and CG14 (*O. glaberrima*), three quantitative agronomic traits, Heading beginning (H), Tiller number at maturity (T) and 1000 Grain Weight (1000GW), were targeted. Bulks of individuals with high and low values of each trait were selected based upon phenotypic measurements. These bulks were sequenced to identify all polymorphisms that might be associated with the trait, differing between the bulks. This provided candidate genes and markers for validation and use in breeding.

P0946: Rice

Restructuration of the Pangenome and Massive Loss of Genes during Domestication: The Case of the African Rice

Cécile Monat[1], Christine Tranchant-Dubreuil[2], Stefan Engelen[3], Karine Labadie[3], Corinne Cruaud[3], Emmanuel Paradis[4], Ndomassi Tando[2] and **François Sabot**[2], (1)Leibniz Institute of Plant Genetics and Crop Plant Research, Stadt Seeland, Germany, (2)IRD - UMR DIADE, Montpellier, France, (3)CEA - Genoscope, Evry, France, (4)ISE-M, Montpellier, France

Pangenome theory implies that individuals from a given group/species share only a given part of their genome (core-genome), the remaining part being the dispensable one.

Domestication process implies a small number of founder individuals, and thus a large core-genome compared to dispensable at the first steps of domestication.

We sequenced at high depth 180 cultivated African rice *Oryza glaberrima* and of 86 wild relatives *O. barthii*, and using a map-the-reassemble approach, we identifed the core and dispensable genome from each species. The wild Pangenome is of 1.6Gb, while the cultivated only 1.4Gb large, and each individuals harbor a quite large fraction of specific sequence compared to another one. After comparing them, we shown that the cultivated species has a smaller core-genome than the wild one, as well as an expected smaller dispensable one. Tis unexpected output however replaces in perspective the inadequacy of cultivated crops to wilderness.

P0947: Rice

Genetic Variation Associated with Iron Toxicity Tolerance in West African Rice Accessions

Giovanni Melandri[1], Juan David Arbelaez[1], M Sikirou[2], K.a. Konaté[3], A.T. Maji[4], A.S. Ngaujah[5], I. Akintayo[6], Vishnu Govindaraj[1], Yuxin Shi[1], Francisco Agosto Perez[1], Venuprasad Ramaiah[7] and Susan McCouch[1], (1)Cornell University, Ithaca, NY, (2)National university of Agriculture, Cotonou, Benin, (3)INERA, Ouagadougou, Burkina Faso, (4)NCRI, Badeggi, Nigeria, (5)SLARI, Freetown, Sierra Leone, (6)CARI, Suakoko, Liberia, (7)Africa Rice Center, Ibadan, Nigeria

Iron (Fe) toxicity associated with acid soils is a major constraint to rice production in West Africa. Previous work has documented variation for Fe toxicity tolerance in rice, with the highest tolerance reported in African rice, *Oryza glaberrima,* and moderate levels in *Oryza sativa*. The breeding program at AfricaRice utilizes *O. sativa*, *O. glaberrima* and *O. sativa x O. glaberrima* populations. Here we report the results of a genome wide association study (GWAS) based on 296 diverse *O. sativa* accessions and a genotyping dataset consisting of 700,000 SNPs. The accessions were evaluated in naturally Fe-toxic soils in Nigeria, Liberia, Burkina Faso and Sierra Leone and under control conditions at IITA (Nigeria). Five traits associated with Fe toxicity response were evaluated: leaf bronzing, days to flowering, plant height, panicle number and grain yield. To examine the genomic variation associated with the QTL regions in both *O. sativa* and *O. glaberrima*, a set of 33 breeding lines from AfricaRice widely used as donors for Fe toxicity tolerance were re-sequenced (~7x genome coverage) and aligned to both the Nipponbare (*O. sativa*) and the CG14 (*O. glaberrima*) reference genomes. As these reference genomes represent independent *de novo* assemblies, there is no common set of coordinates or annotations linking them, making it difficult to identify orthologous variants, genes and QTLs. Comparative analysis is further complicated by structural variation. Efforts are underway to identify the locations of QTLs and informative variants in both genomes, and to analyze the degree of genomic variation/conservation.

P0948: Rice

Introgression of *pi21* By Marker-Assisted Backcrossing Confers Resistance to a Wider Range of Philippine Blast Isolates of Select Rice Varieties

Rosalyn A. Shim[1], Vincent P. Reyes[2], Marilyn M. Del Valle[2], Ruby S. Lapis[2], Kshirod K. Jena[2], Kazuyuki Doi[3] and Motoyuki Ashikari[3], (1)Texas Tech Unviersity, Lubbock, TX, (2)International Rice Research Institute, Los Banos, Laguna, Philippines, (3)Nagoya University, Nagoya, Japan

Blast caused by *Magnaporthe grisea* is a destructive disease of rice that can cause yield losses of up to 100%. The recessive allele *pi21*, is a quantitative blast resistance gene that was isolated from chromosome 4 of the *japonica* rice cultivar Owarihatamochi. To date, *pi21* has only been used to improve blast resistance in *japonica* rice cultivars and has been screened for resistance only against Japanese blast isolates. In the present study, the disease response of the upland rice cultivar Sensho carrying the *pi21* allele to individual or composites of blast isolates from the Philippines was assessed under glasshouse and field conditions. The *pi21* allele was then introgressed by marker-assisted backcrossing from Sensho to four *indica* rice cultivars with unknown or putative blast resistance genes to determine how *pi21* can complement the action of these genes when challenged with different blast isolates from the Philippines. Advance generation lines having the *pi21* in the genetic background of the four *indica* cultivars exhibited resistance to a wider range of Philippine blast isolates compared to the recurrent parents. Agronomic evaluation and grain quality testing of the advance generation lines suggest that *pi21* does not cause penalty in yield components and grain characteristics. Our results show that *pi21* can be used in breeding programs that targets broader spectrum and more durable resistance of *indica* rice cultivars against blast pathotypes in the Philippines.

P0949: Rice

Re-Sequencing and Genotyping of Vietnam's Native Rice Varieties

Bruno A. Santos[1], Jose De Vega[2], Tran Dang Khanh[3] and Mario J Caccamo[1], (1)NIAB, Cambridge, United Kingdom, (2)Earlham Institute, Norwich, United Kingdom, (3)AGI, Hà Nội, Viet Nam

Rice (*Oryza sativa*) is the staple food for half of the world's population and the second most produced staple behind maize. However current rice yield increases are below the pace of population growth. In addition climate change and loss of arable land put food supply in the future at risk. In order to mitigate these challenges genetic gains in rice improvement are needed. One way to achieve this is to mine current allelic diversity within *O. sativa* that can be used to advance rice breeding. In this project a panel of 619 rice lines from Vietnam's national seedbank have been sequenced using low-pass coverage whole genome sequencing (WGS) at the Earlham Institute, UK in collaboration with the Agricultural Genetics Institute (AGI) in Vietnam. The samples cover ca. 10% of the seedbank and have been selected focused on a number of key agronomical traits such as salt tolerance and disease resistance. The generated sequencing data has been analysed using a bioinformatics platform for SNP calling developed at NIAB, UK resulting in more than 5 million high-quality SNP variants. We have characterised the panel structure and identified several subpopulations corresponding to the Indica and Japonica sub-species. In addition to the Vietnamese lines, data from the 3k Rice Genomes Project (IRRI) have been processed using the same pipeline in order to place the genetic diversity contained within the Vietnamese lines into the global collection of rice. Passport and phenotypic data can now be used to further characterise these two races and identify sub-populations within these groups. Finally, the marker density of the panel is suitable for the implementation of genome wide association studies (GWAS).

P0950: Rice

Utilizing Longitudinal Phenotypes to Study and Improve Adaptation to Water Limitation in Rice

Malachy Campbell, Alexandre Grondin, Gota Morota and Harkamal Walia, University of Nebraska Lincoln, Lincoln, NE

Rice is arguably the most important grain crop, and provides a major source of food and economic security for more than 3.5 billion people worldwide. Adequate water availability is essential for proper vegetative growth and grain development. Improving drought tolerance in rice is hindered by the polygenetic nature of the physiological traits underlying adaptation to limited water. Recent advances in high-throughput phenotyping and genotyping have spurred the development of new tools for discovering and recording traits associated with drought adaptation, and understanding their genetic basis. Image-based phenomics facilitates the non-destructive measurement of traits for large populations in highly controlled greenhouse or field environments. This temporal data can be leveraged to uncover complex physiological responses to water availability. Here, image-based phenomics was used to characterize the morphological and physiological responses to water stress over a period of 20 days for a diverse panel of 357 rice accessions. Three traits (shoot biomass, water use, and water use efficiency) were measured daily for each accession in well-watered and water-limited conditions. We applied genomic best linear unbiased prediction (gBLUP) with random regression models to select accessions with contrasting shoot growth responses, water-use, and water-use efficiency. This study (1) demonstrates that random regression models can be used to accurately predict longitudinal phenotypes in both well-watered and water-limited conditions, (2) identifies several accessions that exhibit superior performance under water-limited conditions, and (3) extends the random regression approach to assess the dynamic genetic basis of temporal responses to water limitation through GWAS.

P0951: Rice

A Case Study on a Novel Dehydration Stress Tolerance Mechanism in Transgressive Recombinant Inbred Lines of IR29 x Pokkali: Phenotypic and Transcriptomic Analyses.

Jake Gendron, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX

Transgressive segregants are progenies of bi-parental cross combinations that exhibit novel characteristics beyond the expected phenotypic range of either parent. In an effort to elucidate the regulatory mechanisms behind transgressive phenotypes for stress tolerance in rice, we have reexamined a population of recombinant inbred lines (RILs) derived from a cross between IR29 (salt-sensitive) x Pokkali (salt-tolerant). We subjected the F8 RIL population of to an extensive stress response evaluation at the physiological and whole plant levels under artificial dehydration with 10% Polyethylene Glycol (PEG). Based on a suite of physiological and whole plant level parameters such as recovery, stomatal conductance, chlorophyll content, leaf drying, rolling and temperature, as well as other biochemical properties such as DPPH+, we found that neither Pokkali nor IR29 exhibit tolerance to PEG-induced dehydration. However, a minority of RILs appeared to have acquired mechanisms for tolerance or have become super-sensitive relative to the parents. We conducted an extensive comparison of the PEG-induced transcriptomes across parents and RILs that include positive and negative transgressives. The nature of genetic network rewiring and interactions that characterize the gain of stress physiological attributes in the positive transgressive RIL relative to the parents will be discussed in context of possible synergistic or antagonistic physiological effects. Our results have also indicated that the acquired novel phenotypes exhibited under dehydration stress are independent of the mechanisms responsible for salinity tolerance in the population. Therefore, transgressive phenotypes are likely to be the result of the stacking of positive and complementary mechanisms from both parents.

P0952: Rice

DNA Methylation and Network Rewiring in Salt-Tolerant Transgressive Segregants from a Biparental Recombinant Inbred Population of Rice

Ai Kitazumi[1], Isaiah Pabuayon[1], Rakesh Kumar Singh[2], Yoshihiro Kawahara[3], Glenn Gregorio[2] and Benildo G. de los Reyes[1], (1)Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, (2)International Rice Research Institute, Los Banos, Philippines, (3)National Institute of Agrobiological Sciences, Tsukuba, Japan

We hypothesized that transgressive salt tolerance in a biparental recombinant inbred population of rice represents a possible late-generation consequence of genomic shock and epigenomic landscape reconfiguration due to recombination, leading to network rewiring. In this study, we examined the genome-wide sequence variation, DNA methylation profiles, and transcriptome compositions across the recombinant inbred lines (F_8-RIL) of IR29 (salt-sensitive) x Pokkali (salt-tolerant) in order to understand how the recombinant genomes and epigenomes are

reconfigured and stabilized after multiple rounds of recombination. Our comparative panel included a representative positive transgressive RIL, another RIL with similar level of salt tolerance as the donor parent, and negative transgressive RILs with worse salt sensitivity than the recipient parent. Results showed the occurrence of significant non-parental sequence variations across the RILs, revealing that recombinants underwent significant modifications in the primary DNA sequences. To examine the possible effects of sequence variation to regulatory networks, global transcriptomic and methylomic profiles were scrutinized for non-parental patterns that may correlate to the novel phenotypes in the transgressive segregants. We found that the positive transgressive RIL was unique relative to its parents and other siblings by virtue of its global hypomethylation signature. As a consequence, the transgressive RIL showed transcriptome patterns that indicate massive gains in constitutive expression of genes that are otherwise stress-inducible in the parents and siblings. Regions of the genome that are populated with MITEs were significantly hypomethylated in the positive transgressive RIL, suggesting a link of the gain in constitutive expression to sequence variation and epigenomic reconfiguration.

P0953: Rice

Salt Stress Affects DNA Methylation and Gene Expression Differently in Salt-Sensitive *versus* Salt-Tolerant Rice Cultivars Agami and M103

Nowlan Freese and Ann Loraine, University of North Carolina Charlotte, Kannapolis, NC

Epigenetic engineering of plants to increase abiotic stress tolerance requires understanding how the epigenome and gene expression interact. We characterized the methylome and transcriptome of rice root and shoot tissue from salt-tolerant (Agami) and salt-sensitive (M103) varieties under abiotic salt stress. RNA-Seq analysis identified more differentially expressed genes under salt stress in the salt-sensitive M103 variety compared to the salt-tolerant Agami. For both varieties, differentially expressed genes were mostly up-regulated in shoots and mostly down-regulated in roots. Through whole genome methylation sequencing we found that both varieties had similar numbers of differentially methylated regions in response to salt stress. Salt stress triggered loss of methylation in both Agami and M103 roots. In contrast, salt stress increased methylation in Agami shoots but decreased it in M103 shoots. Gene Ontology analysis identified an enrichment of stress response genes upregulated in Agami roots exposed to salt when compared to M103 roots. Of 76 rice genes with stress response annotations, 30 were differentially methylated. These results highlight how the salt-tolerant Agami response to salt stress varied from the M103 salt-sensitive variety through hypermethylation of the shoots and a comparably small change in gene expression focused on stress response genes.

P0954: Rice

***Xanthomonas oryzae* Pv. *oryzae* Triggers Transcriptional Activation of Diverse Defense-Related Genes in Rice**

Marjohn C. Niño[1], Franz Marielle Nogoy[1], Me-Sun Kim[1], Jae-Young Song[1], Sothea Ouk[1], Yu Jin Jung[2], Kwon-Kyoo Kang[2] and **Yong-Gu Cho**[1], (1)Chungbuk National University, Cheongju, South Korea, (2)Hankyong National University, Ansung, South Korea

Bacterial blight, caused by *Xanthomonas oryzae* pv. *oryzae*, is a serious disease problem of rice causing damage to rice quality and yield. To understand the transcriptional gene network involved in resistance against *Xoo*, a whole-genome oligonucleotide microarray of two popular *japonica* rice Dongjin and Jinbaek were used to infer transcripts of inducible genes between compatible and incompatible interactions at 48 hour post inoculation. A large number of genes are more evident in the resistant cultivar, which is threefold higher than in susceptible plant. Up-regulation of genes with predicted functions in signaling and transcription signifies orchestration of defense signals and robust cellular reprogramming leading to incompatible interaction. To further identify genes crucial to immunity, 13 *Xoo*-DEGs of different protein class were cloned and overexpressed using CaMV 35S promoter into rice. Most of the overexpression plants displayed improved resistance when screened against *Xoo* Korean race K2. Elevated transcripts levels of several defense-related genes at the downstream of defense signal network also corroborate the phenotype reaction of the transgenic plants. ROS levels continuously magnified after inoculation which indicates robust cellular sensing necessary to initiate cell death. Moreover, expression assays revealed regulation of these genes by cross-communicating signal-transductions pathways mediated by salicylic acid. These collective findings revealed the complexity of key immune signaling conduits critical to mount full defense against *Xoo* in rice.

P0955: Rice

The Discovery of Decoy Domains in NLR Receptors Provides Novel Insight into Plant Immunity and Opens New Perspectives for Plant Protection

Thomas Kroj, INRA, Biology and Genetics of Plant-Pathogen Interactions Laboratory, Montpellier, France

Nucleotide-binding domain and leucine-rich repeat containing proteins (NLRs) are important receptors in plant immunity and allow specific recognition of pathogen effectors. Based on our work on the detection of the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by the rice NLR RGA5, we recently developed the hypothesis that some NLRs recognize effectors by integrated decoy domains This 'integrated decoy model' was further supported by work in other experimental systems and has been widely accepted. Comparative genomic analysis showed that NLRs carrying integrated decoy domains are frequent and widespread. We identified them in 31 land plants, from mosses to angiosperms, and they represent, on average, 7% of the NLRs.

By detailed structure-function analysis we further deciphered the molecular details of the binding of AVR-Pia and AVR1-CO39 to the integrated decoy domain of RGA5, a heavy metal-associated domain most related to the yeast copper chaperon ATX1 (RATX1 domain). This demonstrated that the direct RGA5-RATX1/effector binding is strictly required for effector recognition but only of moderate affinity and acts in concert with the association of the effectors to additional sites in RGA5. This combination of integrated decoy domains with additional independent effector-NLR interactions seems to confer robust effector recognition that is resilient to effector mutations. We will present first results on how knowledge on the molecular details of effector recognition by integrated decoy domains can be exploited for the engineering of the recognition spectrum of NLRs.

P0956: Tomato, Potato, Pepper, and related

Sequencing and *de novo* Assembly of the Genome of *Jaltomata sinuosa*, a Species in the Sister Clade to *Solanum* and *Capsicum*, Allows New Comparative Genomic Analyses in the Solanaceae

Meng Wu[1], Jamie Kostyun[2] and Leonie Moyle[1], (1)Indiana University, Bloomington, IN, (2)University of Vermont, Burlington, VT

Jaltomata—a rapidly evolving plant genus in the *Solanaceae*—shows extensive diversity in fruit and nectar color, and flower size and shape, among its ~80 species. Here we report the whole-genome sequencing and assembly of one representative species (*Jaltomata sinuosa*) in this genus. Combining PacBio long-reads (25 X) and Illumina short-reads (148 X) achieved an assembly of approximately 1.45 Gb, spanning ~96% of the estimated genome (1.5 Gb) with a contig N50 of 365 Kb and scaffold N50 of 398 Kb. Searches for complete plant BUSCO gene groups found 96% of curated single copy orthologs in the assembly, supporting a high level completeness of the genome. A large proportion of genomic regions (80%) were identified to be repetitive sequences. In conjunction with transcriptome data from 11 tissues (Wu et al. 2017), 34725 protein-coding genes were predicted. A comparative phylogenetic analyses with six other sequenced *Solanaceae* species indicates that the relative phylogenetic placement of *Jaltomata* with *Solanum* or *Capsicum* cannot be conclusively resolved (i.e., the three species form a polytomy), instead of placing *Jaltomata* as the sister clade to *Solanum* (Olmstead et al. 2008; Särkinen et al. 2013). We also identified gene family dynamics specific to the *Jaltomata* lineage, including the expansion of some gene families potentially involved in novel reproductive trait development, and the loss of gene families that might have accompanied the loss of self-incompatibility in this clade. This high-quality genome assembly will contribute to further studies of genetic mechanisms underpinning phenotypic diversity in this rapidly radiating group, as well as provide a new point of comparison for broader analyses of genetic and genomic evolution across the *Solanaceae*.

P0957: Tomato, Potato, Pepper, and related

Tomato Bioresources in Japan Based on 'Micro-Tom-Japan' to Accelerate the Fruit Biology

Ken Hoshikawa[1], Naoko Ito[1], Tohru Ariizumi[1], Naoya Fukuda[1], Yoshinori Kanayama[2], Yasutaka Kubo[3], Kentaro Yano[4], Koh Aoki[5] and Hiroshi Ezura[1], (1)University of Tsukuba, Tsukuba, Japan, (2)Tohoku University, Sendai, Japan, (3)Okayama University, Okayama, Japan, (4)Meiji University, Kawasaki, Japan, (5)Osaka Prefecture University, Sakai, Japan

Tomato (*Solanum lycopersicum* L.) is one of the most important vegetables in terms of production and economic value. And tomato is useful for studies of fleshy fruit biology and experimental genomic studies of other *Solanaceae* family. Tomato fruits contain many functional metabolites including carotenoids, vitamin A, vitamin C, and GABA. In the future, bioresource collection and preservation will also become increasingly important from the viewpoint of Nagoya protocol and CBD. We have started the tomato bioresource program since 2007 within the framework of the National BioResource Project (NBRP) in Japan (http://tomato.nbrp.jp/). The major purpose of the NBRP-tomato is to collect, preserve and provide tomato bioresources including major experimental lines, mutant lines, transgenic lines and cDNA collections in the genetic background of 'Micro-Tom-Japan' (TOMJPF00001). More than 16,000 mutant lines have been produced by EMS treatment and gamma-ray irradiation, and mutants with visible phenotypes have been isolated. All of the visible phenotyping data and other associated data in individual mutants were registered in the database 'TOMATOMA' (http://tomatoma.nbrp.jp/). We also have measured metabolic components including amino acid compositions, carotenoid contents and Brix values in mutant fruits, and these data has opened through TOMATOMA. These mutants with phenotypic and the metabolite information will help accelerate tomato fruit researches. As DNA resources, Micro-Tom full-length cDNA sequence and EST are available from database 'KaFTom (http://www.pgb.kazusa.or.jp/kaftom/) and EST database 'MiBASE' (http://www.pgb.kazusa.or.jp/mibase/), respectively. Information on genome structural annotations between Micro-Tom and Heinz 1706 is accessible through the genome browser in 'TOMATOMICS' (http://bioinf.mind.meiji.ac.jp/tomatomics/).

P0958: Tomato, Potato, Pepper, and related

Clarification of the Genome Structure of Micro-Tom, a Model Cultivar of Tomato (*Solanum lycopersicum*)

Hideki Nagasaki[1], Kenta Shirasawa[1], Sachiko Isobe[2], Pierre Baldet[3], Christophe Rothan[4], Mohamed Zouine[5], Koh Aoki[6] and Hideki Hirakawa[7], (1)Kazusa DNA Research Institute, Chiba, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)Functional Genomics of Fruit Development, UMR 1332 – INRA, Villenave d'Ornon, France, (4)INRA UMR, Villenave d'Ornon, France, (5)INPT-ENSAT, Toulouse, France, (6)Osaka Prefecture University, Sakai, Japan, (7)Kazusa DNA Research Institute, Kisarazu, Japan

Micro-Tom is one of the cultivar of tomato (*Solanum lycopersicum*), which is known as a major crop and model plant in Solanaceae. Micro-Tom has phenotypic traits such as *dwarf*, and substantial EMS-mutagenized lines have been reported. There are two Micro-Tom varieties, which are Micro-Tom S9 and Micro-Tom MM, that have been maintained independently in Japan and France. Since the whole genome sequencing of Heinz 1706 had been determined, the genome sequence of Micro-Tom was determined by reference guided assembly using 454 FLX reads. To reveal detailed genome structure of Micro-Tom, we have conducted *de novo* assembly of Micro-Tom S9 by adding Illumina MiSeq reads. We obtained the 69M paired-end reads and 269M mate-pair reads. The total coverage was estimated as 63-fold of the Micro-Tom genome. The reads were assembled by MaSuRCA-2.3.2, and BAC end sequences of Micro-Tom S9 were used for scaffolding by SSPACE v2.0, and finally 2,925 scaffolds were constructed (named SLM_r1.1). On the other hand, the high quality pseudomolecule of Micro-Tom MM (Sol_mic1.0) has been built using the optical mapping of BioNano and long read sequencing of Chromium 10x and PacBio by INRA group of France. The scaffolds of SLM_r1.1 were mapped to Sol_mic1.0 by NUCmer in MUMer3.23, and 640 of them were chosen under the condition of 50% length coverage, and then we built a pseudomolecule of Micro-Tom S9. Furthermore we performed SNP and copy number variation (CNV) analyses between Sol_mic1.0 and the 4 Micro-Tom varieties (FRA, JPN, NIVTS, and USA).

P0959: Tomato, Potato, Pepper, and related

Variability in Fruit Ripening within the European Traditional Pool of Tomato Varieties

Clara Pons[1], Jose Blanca[2], Joaquin Cañizares[2], Peio Ziarsolo[2], Richard Finkers[3], José L Rambla[4], Lorenzo Zacarias[5], Antonio Granell[4] and **Antonio J. Monforte**[4], (1)IBMCP. CSIC-UPV, Valencia, Spain, (2)COMAV-UPV, Valencia, Spain, (3)Wageningen UR Plant Breeding, Wageningen, Netherlands, (4)IBMCP, Valencia, Spain, (5)IATA-CSIC, Paterna, Valencia, Spain

Fruit ripening is a highly coordinated and regulated process that involves changes in color, texture, metabolic composition etc. In general, traditional varieties show rapid ripening and shorter post-harvest life than modern hybrid varieties, most of them carrying the ripening-inhibitor (rin) mutation in heterozygosis in order to slow down ripening.
To evaluate postharvest ripening variability present the traditional European tomato pool, 220 varieties composing the Core Collection of European Traditional tomato, representative for the genotypic and phenotypic diversity available in over 1500 TRADITOM repository collection, were studied. Postharvest response was also analyzed in 13 inbreed lines and 27 commercial hybrids, most of them having rin allele in heterozygosis.
Analysis of firmness and color evolution during fruit ripening indicated that, despite the general trend is to be red and soft, differences in the response of some varieties: some remain green and firm, some green and soft and other red and firm. Indicating that color and firmness could proceed at different speeds according to the genotype. In addition, differences between traditional and commercial tomatoes were also observed.
To explore the molecular genetic basis, haplotypes associated to different ripening behavior groups were analyzed. In addition, expression levels of master ripening regulators and key genes involved in the different aspects of ripening were analyzed by using a mid-throughput Fluidigm RT-PCR platform.

P0960: Tomato, Potato, Pepper, and related
The Genomics of Tomato Domestication and Beyond – Using Next-Generation Sequencing to Identify Natural Gene Variants
Niels A. Müller[1], Jose M. Jiménez-Gómez[2], Birgit Kersten[1] and Matthias Fladung[1], (1)Thünen Institute of Forest Genetics, Genome Research, Großhansdorf, Germany, (2)Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France
Understanding the genetic and molecular basis of agriculturally important traits is of fundamental biological interest but also crucial for crop improvement by enabling precision breeding. Next generation sequencing (NGS) technologies greatly facilitate the identification of gene variants underlying complex trait variation. Here I present two examples of genomics-enabled rapid identification of tomato domestication genes and discuss the challenges as well as the exciting opportunities of employing similar methods to study natural variation in forest trees. Cultivated species often experience range expansions. Adaptation to the new agricultural environments thus represents an important feature of domestication. We found that the cultivated tomato experienced marked selection for a decelerated circadian rhythm and reduced photoperiod sensitivity. Quantitative trait locus (QTL) mapping in two densely genotyped populations, and QTL-seq by RNA sequencing of phenotypic bulks, precisely defined the loci underlying the two domestication traits. Analysis of re-sequencing data of several hundred tomato accessions pinpointed a polymorphism in the coding sequence of the light signaling gene *EID1* and non-coding regulatory variation in the flowering gene *SP5G* as the cause of the circadian rhythm deceleration and day neutrality, respectively.
The rapid advance of sequencing technologies opens up unprecedented possibilities for studying the genomics of genetically interesting but complex non-model species, including outcrossing forest trees. By NGS of male and female pools of aspen and silver poplar and subsequent genome scans, we demonstrate that sex determination is genomically highly dynamic in *Populus*. Additionally, our results provide an experimental framework for pool-seq experiments in this important genus.

P0961: Tomato, Potato, Pepper, and related
Genetic Lesion of Yellow Fruit Tomato 1 Gene in 5' UTR Region Alters Fruit Firmness in Tomato
Ling Li[1], Weihua Zhao[2], Xuechao Feng[1], Lulu Chen[1], Guojun Feng[3], Lida Zhang[1] and **Lingxia Zhao**[4], (1)Shanghai Jiaotong University, Shanghai, China, (2)Shanghai Jiaotong University, shanghai, China, (3)Heilongjiang University, Harbin, China, (4)Shanghai Jiaotong Univeristy, Shanghai, China
Fruit firmness is a critical trait associated with the production quality and economic benefit in tomato; however, the precise mechanism of the fruit firmness formation has remained elusive. We found the lesion of the YELLOW FRUIT TOMATO 1(YFT1) gene in 5' UTR alters fruit firmness in tomato, which created opportunity to deeply understand formation of fruit firmness without genetic noises.
To determine effects of the YFT1 lesion on the fruit firmness, yft1 mutant and its wild type of cv. M82 were employed to dissect fruit firmness at morphological and anatomical phenotypes, chemical components, and activity of the catalyzing enzymes, expression of the associated genes and the involved metabolic pathways, and analyzed correlations between aspects.
The lesion of YFT1 resulted in the total of 183 differently expressed genes (DEGs) between yft1 and M82, including 50 down-regulated and 133 up-regulated genes at three developmental stages, which mainly influenced transduction of the ethylene and auxin signals, sugar metabolisms and photosynthesis pathway in tomato. Especially, the lesion also altered expression patterns of numerous genes associated with polysaccharides deposited on the cell wall and sugars accumulated in inside cells, and affected activity of the encoded enzymes, so then elevated fruit firmness of yft1 mutant through promoted deposition of polysaccharides and accumulation of sugars in pericarp cells.
We firstly established correlation between YFT1 and fruit firmness in tomato at different level profiles, which would give new insights into to control fruit firmness formation via regulatory network and exploit new functions of the YFT1 gene, and provide important theoretical informative for improvement of tomato quality in the future.
Keywords: Fruit firmness; yft1 mutant; cell wall; microstructure; polysaccharide; soluble sugar; metabolic enzyme; transcriptomic scale

P0962: Tomato, Potato, Pepper, and related
Phenotypic and Genetic Characterization of New Parthenocarpic Tomato Mutants with a Mutation in F-Box Genes
Farida Damayanti[1], Yoshihiro Okabe[1], Tohru Ariizumi[1] and Hiroshi Ezura[2], (1)University of Tsukuba, Tsukuba, Japan, (2)University of Tsukuba, Ibaraki, Japan
Tomato is one of important vegetable crops worldwide and has been used as a model for plant bearing fleshy fruits. Due to global climate change, high temperature becomes a problem in tomato cultivation because tomato fruit-set is highly affected by it. We have developed comprehensive tomato mutant populations generated by gamma-ray irradiation and EMS mutagenesis in cv. Micro-Tom genetic background as

well as its TILLING platform to screen mutants for the gene of interests. A mutant that derived from EMS mutagenesis and mutated on F-box gene has been identified as a parthenocarpic mutant and has higher Brix value compared to its wild-type. Additional F-box gene mutants, which were discovered by TILLING of target F-Box gene, have also shown higher Brix value by a preliminary evaluation of fruits, but have not been evaluated for their potential of parthenocarpy trait. The objectives of this study were: (1) to investigate the stability of parthenocarpy trait and (2) to characterize the mutants for their morphology, phenology and agronomic traits particularly in relation to parthenocarpy trait and fruit quality. The results showed that the percentage of parthenocarpic fruit formation varied among the mutants and significantly different among three cultivation times. Mutants showed distinct phenotypes, particularly in alteration of leaflet-shape and reduced the number of seed/fruit. Allelism test among the genes, gene expression in NILs of commercial lines, and major metabolites profile will be further investigated to evaluate those F-box tomato mutants as breeding material for higher sugar content and parthenocarpy traits.

P0963: Tomato, Potato, Pepper, and related
The Identification of Parthenocarpic Genes in Tomato by the EMS Mutagenesis Approach
Keiichiro Harada, University of Tsukuba, Tsukuba, Japan
Tomato (*Solanum lycopersicum*) is the vegetable crops whose production is one of the highest worldwide, although its stable production remains an important subject, particularly under harsh environmental conditions such as high temperature and humidity in which productivity significantly decreases resulted from reduced fruit set efficiency. Parthenocarpy is the natural and artificial fruit seedless fruit formation in the absence of pollination and merits attention for improving fruit set efficiency since it can induce fruit formation without requirement of pollination or fertilization. However, genetic loci inducing parthenocarpy is quite limited and thus those involved in parthenocarpy are desired to be identified. For this purpose, we took advantage of the EMS mutagenesis approach using dwarf cultivar 'Micro-Tom' and the parthenocarpic mutant had been selected from mutagenized population. This study reports genetic analysis of the mutant as well as its effectiveness by analyzing yield in summer season. To map the parthenocarpic gene, linkage analysis was carried out using F2 mapping lines derived from a cross between 'Micro-Tom' and 'Regina' and DNA markers spanning tomato chromosomes. Further whole genome sequencing of a mutant was determined and identified DNA substitutions caused by the EMS. Moreover, the mutant showed twofold higher yield compared to wild-type in summer cultivation. We here report the progress of genetic mapping analysis.

P0964: Tomato, Potato, Pepper, and related
Targeting Induced Local Lesions in Genomes (TILLING) in Tomato
Helen Tsai, University of California, Davis, Davis, CA
Continuous improvement of crop varieties is important to address the changing biotic and abiotic pressures and the growing world population. Mutant populations are an excellent source of genetic variation for crop varieties and understanding gene function. Thus, efficient creation and screening of such populations is paramount. We produced two mutant populations in Heinz 1706 tomato (*Solanum lycopersicum*), a legacy commercial canning variety. To do so, we used a reverse genetics approach called TILLING (Targeting Induced Local Lesions IN Genomes) that combines the generation of mutant populations and high-throughput identification of mutations in genes of interest. Populations A and B consist of ~4,000 and ~10,000 individuals, respectively. TILLING lines are non-transgenic and display stable inheritance, allowing them to be readily incorporated into breeding programs. We screened a subset of each of population for an average of twelve ~1.5 kb amplicons and identified a range of mutations including missense, truncation, and splice mutant candidates. Screening the entire A population provides 34% and 97% probabilities of getting a KO and deleterious mutation, respectively. For the B population, those percentages are ~56% and >99%, respectively. We are offering our service to screen these populations with turnover rates of 3-6 months. Use of population A is offered without intellectual property (IP) restrictions while population B has reasonable IP restrictions. Upon completion of the screen, M3 seeds will be distributed. For more information, visit our website (http://tilling.ucdavis.edu/).

P0965: Tomato, Potato, Pepper, and related
Development of Male Sterility Lines by using Marker Assisted Backcrossing (MABC) System of Tomato (*Solanum lycopersicum* L.)
Yu Jin Jung[1], Myoung Kwon Kim[2], Illsup Nou[3], Yong-Gu Cho[4] and **Kwon-Kyoo Kang**[1], (1)Hankyong National University, Ansung, South Korea, (2)Tomato Life Science and Research, Cheongju, South Korea, (3)Department of Horticulture, Sunchon National University, Suncheon, Korea, Republic of (South), (4)Chungbuk National University, Cheongju, Korea, Republic of (South)
Backcross breeding is the most commonly used method for incorporating a male sterility (ms10) gene into a tomato cultivar. Linkage between the ms10 gene and undesirable units can persist for many generations of backcrossing. Marker-assisted backcrossing (MABC) along with marker-assisted selection (MAS) contributes immensely to overcome the main limitation of the conventional breeding and accelerates recurrent parent genome (RPG) recovery. The MABC approach was employed to incorporate (a) male sterility gene(s) from the donor parent HAN 1, into the genetic background of Red fruit, a popular high-yielding line that is male fertility line, to develop a male sterility line improved variety. In this perspective, the recurrent parent genome recovery was analyzed in early generations of backcrossing using SNP markers. Out of 375 SNP markers, 70 markers were found polymorphic between the parents, and these markers were used to evaluate the plants in subsequent generations. Background analysis revealed that the extent of RPG recovery ranged from 65.50% to 90.3% and from 82.40% to 96.70% in BC1F1 and BC2F1 generations, respectively. In this study, the recurrent parent genome content in the selected BC2F2 lines ranged from 92.7% to 97.7%. The average proportion of the recurrent parent in the selected improved line was 95.98%. MAS allowed identification of the plants that are more similar to the recurrent parent for the loci evaluated in backcross generations. The application of MAS with the MABC breeding program accelerated the recovery of the RP genome, reducing the number of generations and the time for incorporating male sterility tomato lines.

P0966: Tomato, Potato, Pepper, and related
Yeast Two Hybrid Analysis of DELLA Variants and SlGID1s of Tomato

Kentaro Ezura[1], Yoshihito Shinozaki[1], Camille Bénard[2], Duyen Prodhomme[2], Yves Gibon[2], Hu Jianhong[3], Tai-Ping Sun[3], Hiroshi Ezura[1] and Tohru Ariizumi[1], (1)University of Tsukuba, Ibaraki, Japan, (2)INRA, Villenave d'Ornon, France, (3)Duke University, Durham, NC

Gibberellin (GA) responses are triggered with GA-mediated recognition of a negative regulatory protein DELLA by the GA receptor *GIBBERELIN INSENSITIVE DWARF1* (*GID1*) and subsequent degradation of DELLA through ubiquitin-proteasome pathway. A naturally found *S. lycopersicum DELLA* (*SlDELLA*), *procera* (*pro*), shows constitute GA responses, inducing stem elongation and seedless fruit development called parthenocarpy. This study newly found a second allele of *procera*, *pro-2*, from a comprehensive EMS mutagenesis population of Micro-Tom, a dwarf cultivar of tomato and confirmed the presence of single base substitution within SAW domain in SlDELLA protein. The *pro-2* showed intermediate stem elongation phenotype and moderate parthenocarpic fruit formation. Genetic allelism tests coupled with complementation experiment, it was demonstrated that *pro-2* is a weaker allele of *SlDELLA* gene. We explored the role of SlDELLA protein in GA signaling through yeast two hybrid (Y2H) assay using SlDELLA and its variants as well as receptor proteins. In this presentation, we report the possible role of SAW domain which might suppress repression activity of SlDELLA.

P0967: Tomato, Potato, Pepper, and related

A Novel Tomato Class I Small SlHSP17.8 Gene is Regulated by SlMADS-RIN Protein in an Ethylene-Dependent Manner

Rakesh K. Upadhyay and Autar K. Mattoo, USDA ARS NEA BARC SASL, Beltsville, MD

Heat shock proteins (HSPs) are ubiquitous in nature and highly conserved in living organisms. In addition to their upregulation in response to heat stress, it is now established that some of them are developmentally regulated. In our laboratory, we have been studying ripening-associated regulation of sHSPs genes. Here, we present studies on a previously unidentified small *SlHSP17.8* gene in tomato, which is a member of a clustered and intronless group of chromosome 6-located sHSP proteins. The *SlHSP17.8* gene encodes a protein of 154 amino acids and possesses characteristic domains of other small heat shock proteins. Its expression is low in vegetative tissues as compared to that in the fruit, with expression increasing further upon ripening. Interestingly, *SlHSP17.8* is specifically up regulated at the fruit transition phase from mature green to breaker, staying high at the breaker stage, and mirrors the expression pattern of *SlACS2,* which encodes the rate-limiting enzyme for fruit-ripening hormone ethylene. Alongside, the expression of another ripening regulator gene *SlMADS-RIN* is in sync with these patterns. *SlHSP17.8* expression is bare minimal in tomato ripening mutants (*rin/rin*, *nor/nor* and *Nr/Nr*) as compared to wild type (WT). Based on these findings, it was apparent that ethylene hormone regulates the expression of *SlHSP17.8* transcripts. *In-vitro* ethylene treatment of WT and ethylene-deficient transgenic line (ACS2-antisense) showed differential suppression of *SlHSP17.8* in both genotypes indicating a dose dependent regulation. *In-silico* promoter studies of *SlHSP17.8* revealed presence of 'CArG' box *cis*-elements that recognize *SlMADS-RIN* protein in many *SlMADS-RIN*-targeted genes, including other sHSP genes. Chromatin immunoprecipitation studies confirmed SlMADS-RIN protein binding to specific 'CArG' motifs present in the *SlHSP17.8* promoter. These data establish SlMADS-RIN protein as a transcriptional regulator of *SlHSP17.8* gene in an ethylene-dependent manner and that this regulation is integral to tomato fruit ripening.

P0968: Tomato, Potato, Pepper, and related

Genetic Control of Immature Fruit Color in Pepper and Tomato

Noam Monsonego[1], Yelena Borovsky[1], Vijee Mohan[1], Sara Shabtai[1], Itzahk Kamara[1], Kevin Stoffel[2], Allen Van Deynze[2] and **Ilan Paran**[3], (1)Agricultural Research Organization, Rishon LeZion, Israel, (2)University of California, Davis, CA, (3)Agricultrual Research Organization, Rishon LeZion, Israel

To study the genetic control of immature fruit color in pepper, we conducted a QTL mapping study in a cross of light and dark green-fruited parents. Two major QTLs, *pc1* and *pc10*, that control chlorophyll content by modulation of chloroplast compartment size in a fruit-specific manner were detected in chromosomes 1 and 10, respectively. The pepper homolog of *GOLDEN2-LIKE* transcription factor (*CaGLK2*) that regulates chloroplast development was identified as underlying *pc10* similarly to its tomato ortholog the underlies the *UNIFORM RIPENING* locus. Fine mapping and bulked DNA and RNA-seq analyses revealed the location of the QTL in chromosome 1. Screening of candidate genes allowed the identification of a gene coding for a transcription factor which is disrupted in the light-green pepper parent. Verification of the gene's function was done by generating CRISPR/Cas9 knockout mutants of the orthologues tomato gene. These tomato mutants exhibited reduction of chlorophyll content similarly to the pepper phenotype, indicating functional conservation of the transcription factor in the solanaceae.

P0969: Tomato, Potato, Pepper, and related

Cloning and Characterization of miR319 in Response to Temperature Stress in Tomato

Xiaopu SHI, Nanjing Agricultural University, Nanjing, China and Zhen Wu, Nanjing Agricultural University/ Key Laboratory of Biology and Germplasm Enhancement of Horticultural Crops in East China, Nanjing, China

MicroRNAs (miRNAs) are a class of small non-coding RNAs that play versatile roles in abiotic stress responses. Temperature stress is the greatest impact abiotic stress during plant growth and development. As the vital vegetable of the world□tomato (*Solanum lycopersicum* L.) is often subjected to chilling and heat stress. Our previous high throughput sequencing results showed that miR319 responded to chilling stress dramatically in the chilling-tolerant wild tomato 'LA1777' (*Solanum habrochaites* L.). However, the function and mechanism of miR319 regulate tomato temperature stress tolerance is still unknown. Here, the precursors of miR319 in *Solanum habrochaites* were cloned from 'LA1777' and the putative target genes *ShaTCP3* and *ShaTCP29* were validated by 5'RACE. Five tomatoes cultivars ('LA1777', 'LA2683', 'Pole red siberian', 'LA3475' and 'Stupice') which have different sensitive to chilling or heat stress were selected to explore the expression patterns of miR319 and its target genes under different temperature stress. Together, our study present evidence supporting the involvement of miR319 in tomato chilling- and heat- stress, the work provides a foundation for further study of the regulation of miR319 in the plant response to temperature stress.

P0970: Tomato, Potato, Pepper, and related

QTL Identification for Tomato Heat Tolerance using Seedling Physiological Indexes

Junqin Wen, NANJING Agricultural University, nanjing, China
Tomato is often affected by unusual high temperature during the growing season, resulting in a decline in yield and quality. Therefore, it is urgent to cultivate heat tolerance tomato varieties for improving tomato production. The study aimed to detect quantitative trait loci (QTL) associated to the heat tolerance physiological indexes of tomato seedling stage, which could lay the foundation for tomato molecular assisted selection (MAS) breeding. In this study, the genetic map was constructed using 144 F2 plants which derived from a cross between the thermo-cultivated tomato 'LA1698' and the heat resistant wild tomato 'LA2093'. The chlorophyll content (Chl), maximum photochemical quantum efficiency (Fv/Fm) of PSII(Photosystem II), electrical conductivity (EC) had been measured and located in QTLs using inclusive composite interval mapping (ICIM). In study, a genetic map was constructed with 70 simple-sequence repeat (SSR) and 34 insertion-deletion (InDel) markers, covering 12 chromosomes of tomato and total distance of 1751.82 cM with an average interval of 16.84 cM. A significant QTL associated with Chl was detected on chromosome 1 and explained 10.94% of phenotypic variation. A QTL was found on chromosome 2 with Fv/Fm and phenotypic variation was 8.84% .When EC was selected, a QTL was detected on chromosome 3, which accounted for 9.56% of phenotypic variation in the population. Three indexes were mapped to three QTLs for heat tolerance. Among them, HT-1 had the highest phenotypic variation, HT-3 additive effect was positive and the rest was negative. As the three indicators were closely related to heat resistance, the detected QTL could be applied to tomato heat tolerance.

P0971: Tomato, Potato, Pepper, and related

Characterization of Facultative Parthenocarpy (*fap*) Mutant isolated from Micro-Tom Tomato Mutant Populations under High Temperature Condition

Satoshi Fujita, Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan, Ken Hoshikawa, University of Tsukuba, Tsukuba, Japan and Hiroshi Ezura, University of Tsukuba, Ibaraki, Japan
Intergovernmental Panel on Climate Change reported that the temperature of global surface will get 0.2ºC forward 10 years. The high temperature will be major problem on food production in the future. In tomato heat stress reduces the pollen fertility and seriously affects the fruit set.
In our previously research, facultative parthenocarpy tomato mutant (*fap*) was isolated from over 4000 Micro-Tom mutant populations generated by EMS treatment. *fap* mutant has significantly higher fruit set and yield than that of WT in summer cultivation (3 times cultivations were conducted in the summers of 2013 to 2015), but in control condition (day/night temperature 25□C/25□C in cultivation room), *fap* mutant showed fruit set with seeds as WT.
The purpose of this study is to identify the detail condition for inducing parthenocarpy phenotype in the *fap* mutant. *fap* mutant was cultivated throughout the year, and measured the temperature and humidity at greenhouse. In the 1st cultivation (January 1 to April 4: day/night temperature 32.5□C/16.2□C, mean of humidity 42.7%), both of WT and *fap* mutant showed over 97% of fruit set with seeds. On the other hand, in the 2nd cultivation (April 11 to July 21: day/night temperature 32.9□C/17.9□C, mean of humidity 53.2%), *fap* and WT produced parthenocarpic fruits. Parthenocarpic fruit set in *fap* mutant was about 30% higher than that in WT. In conclusion, parthenocarpy phenotype in the *fap* mutant is likely to be induced easily by high humidity than WT even under high temperature where WT cannot produce fruits.

P0972: Tomato, Potato, Pepper, and related

Omics-Based Evaluation of Disease Resistance Inducers in Plant

Yuji Sawada[1], Shigemi Seo[2], Muneo Sato[1], Mami Okamoto[1], Naomi Seo[3], Takeru Itabashi[3], Masaaki Osaka[3] and Masami Yokota Hirai[4], (1)RIKEN Center for Sustainable Resource Science, Yokohama, Japan, (2)Institute of Agrobiological Sciences, National Agriculture and Food Research Organization, Tsukuba, Japan, (3)Miyagi Prefectural Institute of Agriculture and Horticulture, Natori, Japan, (4)RIKEN CSRS, Yokohama, Japan
Enhancement of plant resistance to pathogens will contribute to high quality crop yield. Prior treatment of plants with disease resistance inducers is a useful method for activation of the natural defense of plants and has been attracting attention as environment-friendly pesticide. In our previous study, histidine was identified as a significant disease resistance inducer (Seo et al., Plant and Cell Physiology 2016). In next step, multiple field tests with multiple disease resistance inducers are essential for the evaluation. However, such field tests for evaluation often costs a lot and takes time, *e.g.* 1 kg of compounds and 1 year. In this study, we established an evaluation method based on the omics profiles in tomato and strawberry with three types of disease resistance inducers, HIS, prohydroxyjasmone, and probenazole, which are known to activate the defense-related pathways, ethylene, jasmonate and salicylate, respectively. For comparison analysis among inducers, more than one hundred metabolite profiles were measured by using liquid chromatography coupled with tandem quadrupole mass spectrometer. The metabolic profiles were successfully summarized by machine learning models (lasso and elasticnet). Based on the models, key metabolites were predicted for each disease resistance inducers. The key metabolite profiles can be used for defense pathway specific evaluations. Using this method, multiple compounds and multiple plants were successfully evaluated by using only 1 g of each candidate compound within a few weeks. In further plan, we will collect RNA sequence data of tomato samples and the results will be analyzed by machine learning models.

P0973: Tomato, Potato, Pepper, and related

Induction of Broad Range Disease Resistance in Plants using Mutagenesis

Adam Mott[1], Michael Pautler[2], **Ingrid Giles**[2], Darrell Desveaux[1], David Guttman[1] and Daryl J. Somers[2], (1)University of Toronto, Toronto, ON, Canada, (2)Vineland Research and Innovation Centre, Vineland Station, ON, Canada
The level of plant innate immunity that results from perception of microbe-associated molecular patterns (MAMPs) such as ‘flagellin22’ (flg22) by receptors is commonly termed pattern-triggered immunity (PTI). Traditional plant breeding has focused on introducing R-genes into susceptible cultivars as this can result in excellent protection of the new variety; however, this protection is often race-specific or pathogen-specific. Targeting PTI for crop protection is attractive because it promises a more durable and broad-spectrum protection. A previously uncharacterized family of 6 genes encoding immune receptors in *Arabidopsis* was discovered in a screen of insertional knockouts in 169 candidate immune receptors. Knock-out mutants in four of these genes show improved disease resistance against bacteria and oomycetes and are thus termed broad-range resistance (BRR) genes. Translation of this discovery into tomato, pepper and cucumber was initiated by

identifying BRR gene knock-outs in EMS mutant populations of these species. A diagnostic, *in vitro* test for PTI was adopted that measures the peroxidase enzyme activity from leaf discs that are challenged with MAMPs. Tomato homozygous mutants in two genes showed increased activity when challenged with flg22 relative to homozygous wild type segregants. Further, these same mutant plants were challenged with *Pseudomonas syringae* DC3000 to test for pathogen resistance. Homozygous mutants in BRR5 and BRR6 showed 2-fold and 3-fold reduction in bacterial colonization 3 days post inoculation respectively. The research is continuing by examining similar mutant alleles in pepper and cucumber as well as stacking BRR gene mutations to develop higher levels of broad range disease resistance.

P0974: Tomato, Potato, Pepper, and related

Elucidating the Role of MORC1 Protein Interactors during Plant Immunity against *Phytophthora spp*.

Natasha Jackson[1], G. Xu[1], J.C. Nam[2], H.G. Kang[2], D.F. Klessig[3] and P. Manosalva[1], (1)University of California, Riverside, Riverside, CA, (2)Texas State University, San Marcos, TX, (3)Boyce Thompson Institute for Plant Research, Ithaca, CA

Microrchidia (MORC) proteins are a subset of the GHKL ATPase superfamily, containing GHKL and S5 domains that form a catalytically active ATPase module. Proteins containing this GHKL ATPase motif play roles in chromatin remodeling, heat shock responses, signal transduction, and DNA mismatch repair. MORC proteins have been described as components involved in the RNA-directed DNA methylation pathway and chromatin remodeling. Recently, we have found that MORC1 is required for plant immunity against the root rot pathogen, *Phytophthora cinnamomi*. Previously, we reported that MORC1 regulates cell death and plant immunity against *P. infestans* in a species-specific manner behaving as a positive regulator in Arabidopsis and potato and as a negative regulator in tomato and *Nicotiana benthamiana*. We mapped this antagonistic phenotype to the C-terminal region of these MORC1 proteins suggesting that the species-specific effects on resistance are due to how and to whom these MORC1 proteins interact with (positive and negative regulators) at their C-terminal regions. We have identified two proteins that differentially interact with the C-terminal region of the potato and tomato MORC1 proteins. Our results have shown that silencing the *N. benthamiana* homolog of the potato MORC1-Interacting Protein (a transcription factor), compromised the cell death induced by INF1, the major secreted elicitin of *P. infestans*. Furthermore, silencing this transcription factor also increased susceptibility to *P. infestans* and *P. cinnamomi* in *N. benthamiana*. Altogether, our results suggest that this transcription factor also acts as positive regulator of cell death and plant immunity against these devastating oomycete pathogens.

P0975: Tomato, Potato, Pepper, and related

Rpi-amr1*, a Novel Class of Solanum Resistance Genes against *Phytophthora infestans

Kamil Witek, Hari S Karki, Florian Jupe, Agnieszka I. Witek and Jonathan DG Jones, The Sainsbury Laboratory, Norwich, United Kingdom

R gene enrichment and long-read sequencing (SMRT-RenSeq) is a valid tool for identification and cloning of novel functional NLR-type disease resistance genes in plants. Here we utilize SMRT-RenSeq to clone a novel *Resistance to Phytophthora infestans (Rpi)* gene from the wild *Solanum americanum*. Our approach is reference free, and allele mining identified wide distribution of this new gene.
We identified a hotspot for *P. infestans* resistance in numerous *S. americanum* accessions at the distal end of the short arm of Chr 11. We combined bulked segregant analysis and SMRT-RenSeq to clone *Rpi-amr1e* and show that it confers strong resistance against multiple isolates of *P. infestans* in cultivated potato. The gene encodes a typical coiled-coil (CC) NLR protein; however, it belongs to a previously uncharacterized class of CNL genes. In collaboration with Chih-hang Wu in the Kamoun lab, we found that its function is dependent on the helper NLR NRC. We used SMRT RenSeq and association genomics to clone functional *Rpi-amr1e* alleles from several *S. americanum* accessions. Despite 80-90% amino acid identity between these paralogs, they still confer resistance to *P. infestans*. Moreover, employing targeted enrichment-based genotyping, we found that in one of the lines of *S. americanum, Rpi-amr1e* had been translocated to Chr 1.

P0976: Tomato, Potato, Pepper, and related

Re-Analysis of Long Non-Coding RNAs and Prediction of circRNAs Reveal their Novel Roles in Susceptible Tomato Following TYLCV Infection

Jinyan Wang, Jiangsu Acedemy of Agricultural Science, riverside, CA and Baolong Zhang, Jiangsu Acedemy of Agricultural Science, Nanjing, China

Long Noncoding-RNAs (LncRNAs) are known to be involved in some biological processes, but their roles in plant-virus interactions remain largely unexplored. Circular RNAs (circRNAs) have been identified in animals, but litter work has been done on them in plant, especially in the tomato-tomato yellow leaf curl virus (TYLCV) interaction. In this study, pair-end strand-specific RNA sequencing of trancripts from the susceptible tomato line JS-CT-9210 with or without TYLCV inoculation was performed using ribo-zero rRNA removal library method. A total of 2056 lncRNAs including 1767 long intergenic non-coding RNA (lincRNAs) and 289 long non-coding natural antisense transcripts (lncNATs) were obtained. The expression pattern in lncRNAs was similar in susceptible tomatoes between control check (CK) and TYLCV samples. Our analysis showed that lncRNAs might be involved in plant hormone signaling, protein processing in the endoplasmic reticulum, RNA transport, ribosome function, photosynthesis, glulathione metabolism, and plant-pathogen interactions. The silencing of the lncRNA S-slylnc0957 can increase resistance to TYLCV compared with the tobacco rattle virus (TRV) control in susceptible tomato plants. Moreover, we identified 184 circRNAs candidates using the CIRI software, and 32 and 83 circRNAs were specifically expressed in the CK and TYLCV samples, respectively. Most of these circRNAs were derived from exons (62%). We validated the circRNAs using divergent primers by PCR and Sanger sequencing, and found that the expression of exonic circRNAs was correlated with parent genes. Silencing of circRNAs parent genes can decrease TYLCV virus accumulation. In this study, we identified noval lncRNAs and circRNAs using bioinformatics approaches, and showed that the function as susceptible regulators involved in TYLCV infection. Moreover, expression patterns of lncRNAs in susceptible tomato plants were different from resistant tomato plants, and exnoic circRNAs expression positively associated with their respective parent genes. These results build on previous findings and shed new light on the function of lncRNAs and circRNAs in a susceptible tomato breeding line following TYLCV infection.

P0977: Tomato, Potato, Pepper, and related

Identification of Single Nucleotide Polymorphisms on the *Ty-2* Locus to Develop a Gene-Based Marker for TYLCV Resistance in Tomato (*Solanum lycopersicum*)

Sumin Ahn, Jaebeen Lee and **Sung-Chur Sim**, Sejong University, Seoul, South Korea

Tomato yellow leaf curl virus (TYLCV) is a devastating disease in production of tomato (*Solanum lycopersicum*) worldwide. Several TYLCV resistance loci (*Ty*-1 to *Ty*-5) have been previously identified using wild species including *S. chilense*. Of these loci, we investigated sequence variations of the *Ty*-2 locus to develop a gene-based marker for marker-assisted selection in tomato breeding programs. The genomic sequence of *Ty*-2 locus (2.5 kb), which is located on chromosome 11, was obtained from the tomato reference genome assembly v2.50. Two sets of primers were designed to amplify four exon sequences of the *Ty-2* locus in the TYLCV resistant (TB47 and TB50) and susceptible (TB48 and TB49) varieties. The resulting amplicons of these varieties were sequenced using the Sanger method. A total of 12 single nucleotide polymorphisms (SNPs) and one insertion/deletion (InDel) were identified between resistant and susceptible varieties. Of them, three SNPs were found on two exons (1 and 4), while the other SNPs and an InDel were distributed across introns. All three SNPs on exons were synonymous and two were located close to intron/exon boundaries, suggesting that the two SNPs may be responsible for alternative splicing in the *Ty-2* locus. These results will be useful to develop a molecular tool that accelerates tomato breeding for improving TYLCV resistance.

P0978: Tomato, Potato, Pepper, and related

Towards the Design of a Pipeline for the Rapid Generation of Sources of Resistance to Viruses in Tomato

Carlos García-Almodovar[1], Pascual Rodríguez-Sepúlveda[1], Raquel N. Sempere[2], Luís Rodríguez-Moreno[2], Blanca Gosalvez[2], Jesús Agüero[1], **Pau Bretó**[1] and Miguel A. Aranda[2], (1)Abiopep S.L., Murcia, Spain, (2)CEBAS-CSIC, Murcia, Spain

Emerging plant diseases caused by viruses are becoming more frequent in the current scenario of global change. Resistance traits have been identified for a number of viruses and crop species, and cultivars with varying degrees of resistance have been released. However, availability of sufficient sources of resistance and resistance durability are aspects of serious concern to producers, breeders and pathologists. It is therefore necessary to pursue alternative approaches to identify new sources of resistance to viruses in crop species. Taking advantage of the rapid development of genome editing tools, we are working on the design and implementation of a pipeline for the efficient identification of genetic targets whose edition may result in loss-of-susceptibility to plant viruses; our hypothesis proposes that host proteins interacting with viral proteins are likely used by the virus for its own multiplication or transport, and modification or loss of one of these host proteins may result in loss-of-susceptibility. Whitin this conceptual framework, we have identified tomato proteins which interact with *Pepino mosaic virus* proteins. Virus Induced Gene Silencing (VIGS) has shown that at least some of these host proteins are required for full viral infection. Mutants in the corresponding genes have been identified within a TILLING platform, with limited succes. Therefore, genome editing using CRISPR/Cas9 has been set up in our team for tomato, and is under way for melon. Mutants in candidate genes have been generated, and are being characterised phenotypically.

P0979: Tomato, Potato, Pepper, and related

Down the Road to a Pan-Genome Model for Potato

Maria Kyriakidou[1], **David Ellis**[2] and Noelle Anglin[2], (1)McGill University, Montreal, QC, Canada, (2)International Potato Center, Lima, Peru

A reference sequence based on one individual representing a species' genome, is unable to capture the genetic variability among the organisms of that species. A great number of genes affected by Copy Number Variations (CNVs) contribute to diversity of many agronomic traits. Potato (*Solanum tuberosum* L.) is an important staple crop with a highly heterozygous and complex genome. Major efforts for potato improvement have been attempted but the expansion of the available genomic and transcriptomic resources is necessary to explore novel traits. Genomic re-sequencing data from six diploid potato landraces was used to identify structural variation compared to the current reference genome. The genomes were assembled by combining *de novo* and reference based methods. The results of a CNV analysis showed that in the majority of the genomes, the number of the genes affected by deletion events was greater than those affected by duplications. Here, we focus on the genomic analysis of chromosome 12 in these six potato accessions as all appeared to have a high number CNVs per Mb. In particular, genes involved in metabolic process of polysaccharides, environmental stress tolerance, and response to disease had increased copy number in these sequenced genomes relative to the reference.

P0980: Tomato, Potato, Pepper, and related

Optimizing Strategies for Genotyping-by-Sequencing (GBS) in Autotetraploid Potato

Sushan Ru, University of Wisconsin-Madison/University of Minnesota, Madison, WI and Jeffrey Endelman, University of Wisconsin, Madison, WI

Genotyping-by-sequencing (GBS) is a flexible approach for identifying genome-wide polymorphisms and genotyping a panel of individuals simultaneously. Although widely used in diploid species, the economics of GBS is more challenging in polyploids due to the high read depth needed to accurately quantify allele dosage. The objectives of this research were to 1) develop and demonstrate the application of a GBS bioinformatics pipeline in autotetraploid potato and 2) assess the accuracy of GBS genotype calls using potato SNP array data. DNA from a panel of 95 elite russet potato clones was sequenced on two lanes of an Illumina HiSeq2000, allowing us to compare results at 96- versus 48-plex. A bioinformatics pipeline was developed to demultiplex and align reads to the potato reference genome, and the GATK software was used for variant discovery and tetraploid genotype calling. With one lane of sequencing data, 118K SNPs were discovered, and this number increased to 150K with both lanes. Hard filtering with GATK-recommended thresholds resulted in 66K and 86K SNPs for one and two lanes, respectively. Genotype calls from the GBS data were validated using a set of markers in common with the potato 12K SNP array. Genotype call accuracy increased with read depth and was lowest for duplex genotypes, as expected from theoretical calculations.

P0981: Tomato, Potato, Pepper, and related

A High Throughput SNP Genotyping Platform for Potato Breeding and Genetics

Jiazheng (John) Yuan[1], Abdelmajid Kassem[1] and Benoit Bizimungu[2], (1)Fayetteville State University, Fayetteville, NC, (2)Agriculture and Agri-Food Canada, Fredericton, NB Canada, Fredericton, NB, Canada
There are many platforms available for crop genotyping and however, the cost and throughput of those platforms can be greatly different. In our Potato Breeding and Genetics laboratory at Agriculture and Agri-Food Canada, a cost-effective and high throughput platform of genotyping for potato genetics and breeding has been developed and proved to be able to efficiently and accurately determine genotypes of potato lines that are segregating for target traits. In this system, TaqMan® (Life Technologies) based SNP markers have been developed and used in marker assisted selection for breeding. These SNP markers for target traits were originally defined based on genome wide association mapping study (GWAS) on advanced breeding potato selections and commercial cultivars that were previously phenotyped for various traits over many years. The system provides the capacity to accurately and efficiently determine the genotypes of 384 individual samples by the regular PCR in a single plate using the endpoint genotyping software built in a realtime PCR machine. The platform uses a small quantity of genomic DNA obtained from simple DNA extraction. A typical bi-allelic SNP marker derived from Illumina 12K SolCAP chip or genomic sequences containing target SNPs can easily be used as a basic blueprint to design the probes and primers for TaqMan® assays. The platform is suitable for the SNP genotyping of a particular trait on a large number of samples, gene fine mapping, and marker assisted selection and at the same time, it requires minimal hands-on experience.

P0982: Tomato, Potato, Pepper, and related
Engineered PPR10 RNA-Binding Protein for Regulated Gene Expression in Potato Amyloplasts
Qiguo Yu, Rutgers University, Piscataway, NJ
Constitutive, high-level expression of transgenes in the chloroplast compromises plant growth and interferes with development. To restrict transgene expression to potato tubers, we constructed a transgenic system, in which a Green Fluorescent Protein (GFP) is expressed from a mutant PPR10 Binding Site (BSgg) in plastids. Because the wild-type potato PPR10 protein does not recognize the mutant binding site, GFP in transplastomic leaves and tubers accumulates to low levels, with no impact on plant productivity. Tuber-specific GFP accumulation will be achieved by expressing the engineered PPR10gg protein from a tuber specific promoter. Data on tuber-specific expression will be reported at the meeting.

P0983: Tomato, Potato, Pepper, and related
QTL Analysis for Late Blight Resistance in an Andean Tetraploid Potato Population
Juan David Santa S.[1], Jhon Alexander Berdugo C.[1], Teresa Mosquera V.[2], Nubia Liliana Cely[1], Mauricio Soto S.[1] and **Carlos H. Galeano**[3], (1)Corpoica, Mosquera, Colombia, (2)Universidad Nacional de Colombia, Bogotá, Bogotá, Colombia, (3)Corpoica, Palmira, Colombia
Late blight caused by *Phytophthora infestans* is the main problem in potato crop in Colombia. In order to contribute for understanding the genetic basis of resistance to *P. infestans* in potato, this research aim was to identify QTL for late blight using the F1 tetraploid population Roja Nariño × 2384 (RN × 2384). The severity and incidence of the *P. infestans* was evaluated in two crop cycles. The parents and the F1 population were genotyped using a 12K SNP chip, where the 45% of the makers were polymorphic. A genetic linkage map with a length of 968.4 cM was constructed with 1,287 SNPs using the software *TetraploidMap* and the reference physical map for potato (PGSC v4.03). The QTL analysis revealed six QTL linked to *P. infestans* on chromosomes 1, 3, 5 and 8. The most important QTL were qrAUDPC-1 and qrAUDPC-3.2. These results may contribute to potato breeding programs, especially in countries where *P. infestans* is an important factor to limit the potato production.

P0984: Tomato, Potato, Pepper, and related
Genome Wide Association Mapping to Uncover the Genetic Architecture of Morphology in Tetraploid Peruvian Native Potato
Laura M Shannon[1], Rene Gomez[2], Julian Soto[2], Noelle Anglin[2], David Ellis[2] and Jeffrey Endelman[3], (1)University of Minnesota, Saint Paul, MN, (2)International Potato Center, Lima, Peru, (3)University of Wisconsin, Madison, WI
Peruvian native potatoes are the most diverse in the world, exhibiting a range of morphologies. Therefore, they provide an ideal population for uncovering the genetic basis of morphological traits in potato. The International Potato Center (CIP) in Lima, Peru, houses the global in trust collection of potato with over 5,500 accessions of cultivated potato conserved in their genebank making it the largest and most diverse collection in the world. We have genotyped the CIP collection with version 2 of the SolCAP SNP array, creating the most comprehensive potato diversity panel to date. The panel contains over 1687 predominantly tetraploid cultivars, with morphological descriptors developed as part of genebank curation. Of the 12K SNPs on the array, accurate tetraploid genotype calls were made for 9292 using the ClusterCall package in R. To identify the genetic basis for the morphological descriptors, we implemented a Genome Wide Association Mapping (GWAS) approach using linear mixed models. The resulting QTL for coloring and patterning across tissues exhibited a range of dominance patterns, including single and multiple allele dominance. Many of the resulting QTL co-localized near two known color loci, *Developer* and *Pigmented Flesh,* on chromosome 10. However, the majority of the QTL are for traits not known to be controlled by either gene. *Developer* is a MYB transcription factor, one of seven in the QTL region. We hypothesize a series of diverged homologous MYBs and associated cis regulatory elements may be the primary determinants of color and patterning across tissues in Peruvian native potatoes.

P0985: Tomato, Potato, Pepper, and related
Investigating Resistance Mechanisms to Colorado Potato Beetle in Diploid Potato
Natalie R Kirkwyland[1], David Douches[2], Joseph Coombs[1], C. Robin Buell[3] and Norma C. Manrique Carpintero[1], (1)Michigan State University, East Lansing, MI, (2)Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, (3)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI

Potato production is hampered by the Colorado potato beetle (*Leptinotarsa decemlineata*) (CPB) and conventional control is stymied by its staggering resistance to almost every class of synthetic insecticide. Host plant resistance offers a valuable range of tools to complement existing management practices but introgression has been fettered by the complexity of polyploid potato genetics. Our goal is to exploit the power of inbreeding and inbred germplasm to investigate mechanisms of innate CPB resistance in diploid potato to further equip the potato community with accessible and exploitable plant host resistance to control this insect pest. We generated a mapping population from the largely homozygous *S. chacoense* lines USDA8330-1 and M6, segregating for accumulation of the novel glycoalkaloids leptine and leptinidine. The F_2 generation demonstrated 1:2:1 segregation for CPB host plant resistance in a replicated field trial in May 2017. To efficiently identify loci underlying the resistant trait in these closely related F_2 lines, we employed a bulk segregant approach coupled to whole genome sequencing (QTL-seq) in tandem with traditional QTL analysis, utilizing 1200 SNPs produced from the V3 22K SNP Illumina array. Although we hypothesize that glycoalkaloids are the primary driver of resistance, preliminary detached leaf bioassays and targeted metabolic analysis indicate more complex chemical interactions are responsible for this trait. As such, integrated RNA-seq and untargeted metabolite profiling will enable expression characterization of candidate genes identified by QTL analysis. This project will facilitate the development of markers to efficiently transfer resistance harbored in wild relatives to cultivated material.

P0986: Tomato, Potato, Pepper, and related

Spatiotemporal Analysis of Potato Hypersensitive Resistance Response to Potato Virus Y: RBOHD Is Required for Successful Virus Arrest

Kristina Gruden[1], Spela Baebler[1], Tjasa Lukan[2], Andrej Blejec[2], Maja Zagorscak[2] and Marusa Pompe Novak[2], (1)National Institute of Biology, Department of Biotechnology and Systems Biology, Ljubljana, Slovenia, (2)National institute of biology, Slovenia

Mechanisms of viral arrest in hypersensitive resistance (HR) response have remained elusive despite intensive ongoing research. We here show that in potato –potato virus Y HR interaction, as shown in some other plant-virus pathosystems, it is not the cell death or physical separation of the virus by callose deposition that prevents virus spread. Additionally, no particular ultrastructural features were detected distinguishing HR and non-HR response. Thus, we carried out detailed study of spatiotemporal transcriptional responses. We have chosen 32 candidate genes for our analysis based on time stamped whole leaf transcriptomic data. These included genes involved in ethylene, jasmonate and salicylate signaling, metabolism of reactive oxygen species, response to redox potential changes and a set of immune signaling actuator genes. Analysis of small sections surrounding the site of viral infection has shown that responses for almost all selected genes are tightly spatiotemporally regulated. Interestingly, response of redox state related genes was showing spatiotemporal response that differed between resistant and susceptible genotypes. In particular, responses of RBOHD gene are focused on the border region of the lesion and correlate with the expression of TRX-H gene known to be involved in regulation of SA signaling. Based on these results we hypotesised that RBOHD is essential for signaling leading to successful arrest of the virus in HR response of potato. We have constructed transgenic lines with suppressed RBOHD gene activity. The virus can spread systemically in those plants, breaking the HR response, thus validating our hypothesis.

P0987: Tomato, Potato, Pepper, and related

Genomewide Chromosomal Rearrangements and Impact of Positional Effect of Associated SNPs for Capsaicinoids and Fruit Weight in various *Capsicum spp.*

Umesh K. Reddy[1], Padma Nimmakayala[1], Lakshmi V. Abburi[1], Carlos Ortiz[1], Thangasamy Saminathan[1], Suresh B. Alaparthi[1], Brittany Davenport[1], Marjan Nadimi[1], Gopinath Vajja[1], Tolulope Akinmoju[1], Tosin Akinspe[1], Seungill Kim[2], Doil Choi[3] and John Stommel[4], (1)Department of Biology, West Virginia State University, Institute, WV, (2)Seoul National University, Seoul, Korea, Republic of (South), (3)Dept of Plant Science Seoul National University, Seoul, Korea, Republic of (South), (4)Genetic Improvement of Fruits and Vegetables Laboratory, Beltsville, MD

The genus *Capsicum* originated in Bolivia and consists of 25 to 30 species; five are domesticated: *C. annuum, C. baccatum, C. chinense, C. frutescens,* and *C. pubescens*. We used a thousand accessions representing all the species and performed Genotyping by Sequencing to produce 66,000 SNPs that are mapped to *C. annuum, C. baccatum* and *C. Chinense* genomes. Common SNPs mapped to 3 genomes were insightful to locate chromosomal rearrangements in these three cultivated genomes. PCA and IBS were used in a mixed linear model of traits capsaicin and dihydrocapsaicin content and fruit weight observed for several seasons. Our research focused on identifying SNPs for various traits and SNP positional effect and impact of genome rearrangements that moved associated SNPs to different locations.

P0988: Tomato, Potato, Pepper, and related

New Reference Genome Sequences of Hot Pepper Reveal the Massive Evolution of Plant Disease-Resistance Genes by Retroduplication

Seungill Kim, Seoul National University, Seoul, Korea, Republic of (South), Seon-In Yeom, Gyeongsang National University, Seoul, South Korea, Won-Hee Kang, Institute of Agriculture & Life Science, GNU, Jinju, Gyeongsangnam-do, Korea, Republic of (South) and Doil Choi, Dept of Plant Science Seoul National University, Seoul, Korea, Republic of (South)

Transposable elements are a major evolutionary forces which can cause new genome structure and species diversification. The role of transposable elements in the expansion of nucleotide-binding and leucine-rich-repeat proteins (NLRs), the major disease resistance gene families, has been unexplored in plants. We report two high-quality *de novo* genomes (*Capsicum baccatum* and *C. chinense*) and an improved reference genome (*C. annuum*) for peppers. Dynamic genome rearrangements involving translocations among chromosomes 3, 5 and 9 were detected in comparison between *C. baccatum* and the two other peppers. The amplification of *athila* LTR-retrotransposons, members of the *gypsy* superfamily, led to genome expansion in *C. baccatum*. In-depth genome-wide comparison of genes and repeats unveiled that the copy numbers of NLRs were greatly increased by LTR-retrotransposon-mediated retroduplication. Moreover, retroduplicated NLRs are abundant across the angiosperms and in most cases are lineage-specific. Our study reveals that retroduplication has played key roles for the massive emergence of NLR genes including functional disease-resistance genes in pepper plants.

P0989: Tomato, Potato, Pepper, and related

Utilizing Wild *Capsicum annuum* (Chile Pepper) for Breeding Beet Curly Top Virus Resistance in Cultivated Hot Peppers

Randi C. Jimenez, University of California, Davis, Davis, CA and **Allen Van Deynze**, University of California, Davis, CA

Geminiviruses are the largest family of viruses threatening global vegetable production. Additionally, *Beet curly top virus* (BCTV) is one of the most damaging geminivirus of chili pepper (*Capsicum annuum*) in the United States that can result in yield losses ranging from 20-80%. BCTV is transmitted by leafhoppers (*Circulifer tenellus*) and infect a wide range of plants, such as pepper, bean, sugar beet, tomato, cucurbits and spinach. Both the virus and the insect vector continue to be difficult to control. Our goal is to investigate germplasm sources from landraces collected in Mexico, where virus is prevalent, as well as lines from the literature for resistance to BCTV. To identify sources of resistance, we utilize a rapid *Agrobacterium*-mediated inoculation assay. Several accessions identified as resistant have been crossed into a cultivated, susceptible jalapeño variety to generate and three test populations segregating for BCTV resistance and favorable traits. Wild accessions were preferentially selected based on traits such as seed production, fruit type, determinacy and the ability to cross with other *C. annuum*. These populations are being used to determine the genetics of BCTV resistance in pepper. Our long-term goals are to develop and release pepper breeding lines that combine resistance from wild pepper germplasm to BCTV, as well as to determine the genetic basis of this resistance. Identifying genetic resistance from multiple sources is the key to integrated management programs to protect yield and quality in pepper and other crops.

P0990: Tomato, Potato, Pepper, and related

Investigation of *TSW* Functions for *Tomato Spotted Wilt Virus* Resistance (TSWV) in *Capsicum*

Min-Young Kang, Seoul National University, Seoul, South Korea, Won-Hee Kang, Institute of Agriculture & Life Science, GNU, Jinju, Gyeongsangnam-do, Korea, Republic of (South) and Byoung-Cheorl Kang, Seoul National University, Seoul, Korea, Republic of (South)

Plant viruses cause various types of plant diseases. *Tomato spotted wilt virus* (TSWV) is a destructive pathogen affecting crop production. *Tsw* is a single dominant gene derived from distinct *C. chinense* accessions that controls the hypersensitive response (HR) to most Tomato spotted wilt virus isolates. We identified the *Tsw* gene by a genomics-assisted cloning method. The *Tsw* gene encodes typical coiled-coil (CC), nucleotide-binding (NB-ARC), and leucine-rich (LRR) domain protein. CC domain produced a strong HR without TSWV infection, whereas NB-ARC and NB-ARC-LRR did not induce HR when these domains were transiently expressed in *N. benthamiana*. A strong HR was observed when CC and NB-ARC or NB-ARC-LRR were co-expressed *in trans*. However, CC-NB-ARC did not induce HR *in cis*. indicating that NB-ARC is sufficient to suppress CC-induced HR *in cis* but not *in trans*. To examine which region of the *Tsw* gene is required for triggering HR by TSWV, different domains or domain combinations including NB-ARC, NB-ARC-LRR, or deleted CC-NB-ARC-LRRs were constructed. All combinations produced no HR by NSs of TSWV demonstrating that a full length CC domain is required for triggering HR. The *Tsw* gene is very similar to a *Potyvirus* resistance gene *Pvr4*. Although *Tsw* and *Pvr4* genes have been derived from a common ancestor, these genes recognize totally different viruses. To investigate important domains for recognition of different viruses, we obtained six chimeric genes (*C1-C6*) by the golden gate cloning method. Chimeric genes were transiently coexpressed with avirulence genes of PepMoV and TSWV. Strong HR responses were detected on infiltrated leaves of *C1* and *C4* with the avirulence domains of TSWV and PepMoV, respectively. The *C4* gene recognized both TSWV and PepMoV.

P0991: Tomato, Potato, Pepper, and related

Functional Analysis of the Vacuolar Processing Enzymes Family in Plant Development in Sweetpotato

Jiaojiao Jiang, Dan Su, Yonghua Han and **Zongyun Li**, School of Life Science, Jiangsu Normal University, Xuzhou, China

Vacuolar processing enzymes (VPEs) belongs to cysteine proteases which are responsible for maturation of vacuolar proteins and associate with development and immunity in higher plant. Subsequent studies demonstrated VPEs harbouring Caspase-1-like activity, which is key enzymatic activity for apoptosis in animal. VPEs was found to initial trigger program cell death (PCD) in plant cell and play an important role in plant tolerance to various biotic or abiotic stresses. However, developmental function of VPEs is very limited. Recent studies in tomato demonstrated an additional function of VPEs involved in sucrose accumulation and fruit ripening, suggesting a new function of VPEs during plant development. In this study, we aimed to determine the roles of VPEs in development of sweetpotato. Three potential genes encoded VPEs, namely *Ibvpe1*, *Ibvpe2*, *Ibvpe3*, were characterized from sweetpotato by proteomics and genomic approaches. Expression patterns of these *vpes* were analyzed in tissue-specific and various development stage, and localization of *vpes* was confirmed by confocal analysis. Functions of *vpes* were determined by analyzing *vpe*-overexpressing transgenic plants and loss-of-function mutants. These results suggest that function of *IbVPE* family is important for development of sweetpotato.

P0992: Tomato, Potato, Pepper, and related

Genomics-Assisted Breeding for Sweetpotato Improvement

Xiaofei Zhang, Bode Olukolu, Ken Pecota and Craig Yencho, North Carolina State University, Raleigh, NC

Sweetpotato [*Ipomoea batatas* L. (Lam.)] produces edible tuberous roots. It was originally found in Central America and now becomes one of the most important food crops in the world. Its global production is around 104 million tons each year. Developing superior cultivars with improved yield, disease resistance, nutrient richness, and processing quality is key for boosting sweetpotato production. However, because of its auto-hexaploidy (2n = 6× = 90), high heterozygosity, big genome, and self-incompatibility, developing molecular markers for genetic analysis and molecular breeding has been challenging. In the present study, using genotyping-by-sequencing (GBS) technology, we discovered genome-wide markers from a breeding population with 310 accessions, and performed the proof-of-concept study on genomic selection in sweetpotato. Genome-wide markers, the optimization of the training population and the predictive ability on yield and disease resistance will be reported. Moreover, we obtained genome-wide markers from a population consisting of released cultivars and main parents and advance clones in sweetpotato breeding programs. The relatedness of these sweetpotato germplasm will be presented and its application in sweetpotato improvement will be discussed.

P0993: Tomato, Potato, Pepper, and related

Unstable Allotetraploid Tobacco Genome due to Frequent Homeologous Recombination, Segmental Deletion and Chromosome Loss

Shumin Chen, Feihong Ren and **Hanhui Kuang**, Huazhong Agricultural University, Wuhan, China

All sequenced plant genomes showed whole-genome duplication (WGD), suggesting polyploid history. The major genetic events and their relative importance in shuffling polyploid genomes remain unknown. In this study, a high throughput method was developed to identify loss-of-function alleles for the resistance gene *N*, which provides resistance against the tobacco mosaic virus (TMV) in allotetraploid tobacco. A total of 2,134 loss-of-function alleles of the *N* gene were identified after screening 14 million F_1 hybrids. The loss of *N* gene function was mainly caused by homeologous recombination and loss of the N-containing chromosome. Point mutations and short indels played only limited roles. The frequency of indel incidence in the *N* gene was estimated to be 5.3×10^{-11} per site per generation. Homeologous recombination of the N-containing chromosome occurred only in subtelomeric region, with a frequency of ~1/12,000. On the other hand, we observed no homeologous recombination and a spontaneous segmental deletion frequency of approximately 1/16,000 in a distinct chromosome called the P-chromosome. The frequency of loss of the P-chromosome (~1/13,000) was similar to the frequency of loss for the N-containing chromosome (~1/15,000). Both homeologous recombination and chromosome loss considerably decreased the viability of the mutants. The high mutation rate and decreased mutant viability in allotetraploid tobacco are consistent with the hypothesis of an evolutionary dead end for polyploids. Frequent mutations tend to drive polyploids to extinction unless a novel mutation helps the polyploid to effectively compete with diploids or find a new ecological niche.

P0994: Wheat, Barley, Oat, and related

The International Wheat Genome Sequencing Consortium

The International Wheat Genome Sequencing Consortium[1], **Kellye Eversole**[2], Jane Rogers[3] and Isabelle Caugant[1], (1)IWGSC, Lee's Summit, MO, (2)IWGSC, Bethesda, MD, (3)IWGSC, Cambridge, United Kingdom

Bread wheat, the staple food for 35% of the world's population and the most widely produced crop, is one of the last important crop species to benefit from a comprehensive set of high quality genomic resources. Genomics offers powerful tools for understanding the molecular basis of phenotypic variation as well as accelerating gene cloning, marker assisted selection, and more efficient exploitation of genetic diversity.

The IWGSC, with 1,800 members in 62 countries, is an international, collaborative consortium, established in 2005 by a group of wheat growers, plant scientists, and public and private breeders. The goal of the IWGSC is to produce a high quality, annotated reference genome sequence for bread wheat as a foundation for future research and breeding efforts in wheat improvement.

In 2017, the IWGSC made available to the scientific community the first high quality reference sequence of the bread wheat variety *Chinese Spring* (IWGSC RefSeq v1.0), along with annotation encompassing genes, repeat sequences and non-coding RNAs. Since the release of the annotation in June 2017, the IWGSC has continued to align the reference with additional data and perform whole genome, subgenome, and chromosomal analysis.

An overview of the latest available data will be presented along with an outline of the IWGSC Phase II activities.

P0995: Wheat, Barley, Oat, and related

The Reference Sequence for the Bread Wheat Genome

The International Wheat Genome Sequencing Consortium, IWGSC, Lee's Summit, MO and **Rudi Appels**, Murdoch University, Perth, Australia

The goal of the IWGSC since its inception in 2005 has been the generation of a high quality reference genome sequence for bread wheat that integrates genetic and genomic resources and supports the acceleration of wheat production and improvement to keep pace with projected rises in human population.

The IWGSC RefSeq v1.0 has been assembled from a whole genome shotgun assembly of allohexaploid bread wheat cv. Chinese Spring and integrated with a wealth of community resources including: chromosome survey sequences (IWGSC CSS), chromosome-specific BAC-based physical maps, WGP™ tag sequences and optical maps, POPSEQ genetic maps, Hi-C and radiation hybrid maps.

The assembly represents ~94% of the predicted wheat genome size in large scaffolds (N50 22.8Mb) that are assigned and ordered along the 21 wheat chromosomes. 107,886 high confidence gene models have been annotated and further sets of incompletely supported sets of gene models and pseudogenes identified. The annotated genes have been used to analyze the distribution of homoeologous genes across A,B,D genomes, together with gene duplications and losses that play important roles in wheat evolution. Insights into gene expression and its regulation have been revealed using a transcriptome atlas developed from 850 RNASeq datasets representing all stages of wheat phenological development. With a sequence assembly that now supports the resolution of complex gene families associated with important traits such as yield, grain quality or disease resistance, wheat now has a key resource in place to anchor all QTL knowledge to the reference sequence and stimulate new molecular approaches for the future.

P0996: Wheat, Barley, Oat, and related

Global Crop Diversity Analysis to Explore and Unlock the Genetic Potential of the World's Largest Wheat Germplasm Collections (CIMMYT and ICARDA)

Carolina Paola Sansaloni[1], Jorge Franco[2], Bruno A. Santos[3], Mario J Caccamo[3], Sukhwinder Singh[1], Cesar Daniel Petroli[4], Thomas Payne[1], David Marshall[5], Benjamin Kilian[6], Carolina Saint Pierre[7], Andrzej Kilian[8], Peter Wenzl[9], Mohamed Fawzy Nawar[10], Kai Sonder[11], Jaime Amador Campos Serna[7] and Kevin Pixley[1], (1)CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico, (2)Facultad de Agronomia, UDELAR, Paysandu, Uruguay, (3)NIAB, Cambridge, United Kingdom, (4)CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico, Mexico, (5)The James Hutton Institute, Dundee, United Kingdom, (6)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (7)CIMMYT- International Maize and Wheat Improvement Centre, Mexico, (8)Diversity Arrays Technology Pty Ltd, Canberra, Australia, (9)CIAT, Cali,

Colombia, (10)International Center for Agriculture Research in the Dry Areas (ICARDA), Rabat, Morocco, (11)CIMMYT - International Maize and Wheat Improvement Centre, El Batan, Mexico
Climate change and the rapidly growing human population trigger the demand to explore and unlock unutilized genetic resources for feeding future generations. In wheat, undomesticated wild species, crop wild relatives, and landraces represent sources of new variation for cultivar improvement. These materials provide alternative mechanisms to face disadvantageous conditions, as they have survived extreme environmental challenges and continuous cycles of natural selection. However, their resilience and adaptive capacity mechanisms remain largely untapped and poorly understood. The Seeds of Discovery initiative (SeeD) (http://seedsofdiscovery.org/), a pioneering project led by CIMMYT and aiming to unlock and utilize novel genetic diversity held in genebanks, has characterized, using DArTseq™ technology, more than 100,000 genetic profiles from the two biggest wheat genebanks in the world (~40% of the CIMMYT Germplasm Bank and almost 100% of the ICARDA Germplasm Bank). The global diversity analysis was divided into three biological categories: 4,206 wild relatives (WR) accessions, including 27 species from 55 countries; 20,000 tetraploid (AB genome), including 10 domesticated species from 76 countries; and 60,000 hexaploid (ABD genome) with 7 domesticated species from 82 countries. Our analysis has identified 450K, 130K and 200K high-quality DArTseq SNPs and SilicoDArT (presence/absence variations) markers for WR, tetraploid and hexaploid sets respectively. All markers generated were aligned to the IWGSC v0.4 reference genome (https://urgi.versailles.inra.fr/download/iwgsc/) and against the DArT consensus map (diversityarrays.com). To date this represents the largest crop genotyping effort, generating resources to underpin the breakthroughs necessary to develop the crops of the future.

P0997: Wheat, Barley, Oat, and related

Wisconsin Crop Innovation Center – a Public Resource for Plant Transformation and Editing Research

Heidi F. Kaeppler[1], Michael W. Petersen[2], Brian J. Martinell[2], Amy J. Miyamoto[2], Vai Sa Nee Lor[3] and Shawn Kaeppler[4], (1)Department of Agronomy, Wisconsin Crop Innovation Center, University of Wisconsin, Madison, WI, (2)Wisconsin Crop Innovation Center, University of Wisconsin Madison, Middleton, WI, (3)Wisconsin Crop Innovation Center, University of Wisconsin Madison, Middleton, WI, Middleton, WI, (4)Department of Agronomy and Wisconsin Crop Innovation Center, Madison, WI
Genetic engineering and gene editing systems are critical tools for the advancement of crop functional genomics research and genomics-based crop improvement efforts both in the U.S. and worldwide. Current crop transformation systems are limited, however, by genotype specificity of transformation protocols, high complexity and low efficiency of the processes, variable responses of target tissues, lack of high-throughput procedures, intellectual property-related restrictions, and an overall lack of capacity at the national and international levels. In January of 2017, the Wisconsin Crop Innovation Center opened its doors with the intent to deliver new innovations and processes to public researchers with the aim of reducing the plant transformation bottlenecks that are evident in present day systems. WCIC is working with researchers by providing high capacity transformation services, conducting collaborative research in systems development, and initiating independent research in transformation and gene editing system improvements. Key benefits that a facility like the WCIC brings to the crop genomics research community include high capacity, significant knowledge in transformation system establishment and optimization, rapid plant turnaround, collaborative opportunities, genotype independent transformation in certain species, and new technology development. The WCIC has already added elite soybean and cowpea to its portfolio and is currently working on B73 and ex-PVP maize varieties, potato, dry bean, sorghum, cucurbits, chickpea, tobacco, tobacco chloroplast, switchgrass, cassava, and poplar. In 2018 we hope to engage additional researchers and collaborators that are seeking larger scale transformation and editing projects in a variety of plant species. Our poster will provide research updates and describe the capabilities of the WCIC.

P0998: Wheat, Barley, Oat, and related

Reliable and Efficient High-Throughput Phenotyping to Accelerate Genetic Gains in Norwegian Plant Breeding

Morten Lillemo[1], Jon Arne Dieseth[2], Pål Johan From[1], Ingunn Burud[1], Are Søby Vindfallet[3], José Crossa[4], Osama Alshaykh[5] and Muath Alsheikh[2], (1)Norwegian University of Life Sciences, Aas, Norway, (2)Graminor, Ridabu, Norway, (3)Making View, Hamar, Norway, (4)CIMMYT - International Maize and Wheat Improvement Centre, Mexico, D.F., Mexico, (5)Boston University, Boston, MA
New technologies like genomic selection and high-throughput phenotyping offer possibilities to increase genetic gains in plant breeding through more precise selection and shortening of the breeding cycle. However, considerable research is needed in terms of theoretical developments, statistical modeling and technical solutions to achieve this in practice. Here, we present the recently funded 4-year Norwegian project “Virtual phenomics” (vPheno, NFR 267806, 2017-2021), which is a collaboration between the plant breeding company Graminor, the Norwegian University of Life Sciences and world-leading groups in robotics, image analysis, statistical modeling and data management. In this project, we will develop novel statistical models to extract biologically relevant information from multispectral and hyperspectral images. The work will consist of developing reliable methods for capturing high-resolution images of field plots, and utilizing novel computational solutions to integrate top view images from drones with close-up images from robots to build 3D models that retain the original resolution and multispectral information. Computational algorithms will be used to extract important physical and physiological traits that can be used directly as selection tools in plant breeding. By coupling multispectral data with grain yield and other direct measurements, statistical prediction models will be developed that plant breeders can use in early-generation selection to increase yield gains. User-friendly solutions will be developed through direct involvement of plant breeders in the project. By utilizing virtual reality technology, our ultimate goal is to "take the field to the breeder" and let the plant breeder observe the field plots and associated data through VR goggles.

P0999: Wheat, Barley, Oat, and related

The Genome of *Triticum urartu*, a Progenitor of Wheat a Genome

Hong-Qing Ling, Bin Ma, Xiaoli Shi, Hui Liu, Lingli Dong, Aimin Zhang, Daowen Wang and Chengzhi Liang, Institute of Genetics and Developmental Biology, CAS, Beijing, China

Bread wheat (*Triticum aestivum*, Ta), an allohexaploid species (2n = 6x = 42) containing three homoeologous subgenomes (A, B, and D), is one of the most important global food crops. It supplies approximately 20% of the protein and dietary calories consumed by humans. Bread wheat evolved through two polyploidisation events: the first involved the hybridisation between the ancestral species carrying the A (*T. urartu*) and B (unknown) genomes to form tetraploid wheat; the second involved hybridisation between the tetraploid wheat (AABB) and *Aegilops tauschii* (Aet, DD). The accurate sequencing and assembly of wheat genomes will help to elucidate wheat evolution and genetic variation, and will facilitate the genetic improvement of wheat varieties.
Triticum urartu, a wild diploid wheat, is the progenitor of the A subgenome of tetraploid and hexaploid wheat. Ample genetic studies have shown the value of *T. urartu* for investigating the structure, function, and evolution of polyploid wheat genomes. Here, we report the generation of a high-quality genome sequence of *T. urartu* by combining BAC-by-BAC sequencing, single molecule real-time (SMRT) sequencing, and next-generation mapping (BioNano genome map and 10x Genomics linked reads) technologies. We produced seven chromosome-scale pseudomolecules that spanned 4,666 Mb and annotated 37,516 high confidence and 3,991 low confidence protein-coding genes. By comparing collinear segments between *T. urartu* and its grass relatives rice, sorghum, and *Brachypodium*, we propose an evolution model of *T. urartu* chromosomes, and found that *T. urartu* and *Brachypodium* were independently evolved from the grass ancestor with 12 chromosomes. Furthermore, the ancient genome duplications, which are well maintained in rice, sorghum, and *Brachypodium,* were strongly corrupted in *T. urartu* because of extensive amplifications of transposable elements and widespread gene loss. Overall, the *T. urartu* genome sequence described here provides a valuable reference for systematic studies of Triticeae genomes and for genetic improvement of wheat.

P1000: Wheat, Barley, Oat, and related
A 16 Founder Wheat MAGIC Population for QTL Mapping
Nicholas Fradgley[1], Alison R Bentley[1], Keith A. Gardner[1], Phil Howell[1], Ian J Mackay[1], Mike Scott[2], Richard Mott[2] and James Cockram[1], (1)NIAB, Cambridge, United Kingdom, (2)UCL, London, United Kingdom
Development of experimental mapping populations enables dissection of complex traits in crops and underpins modern plant breeding. Plant Multi-parent Advanced Generation Inter-Cross (MAGIC) populations have now been constructed in many crop species. The advantages of MAGIC include greater genetic diversity captured from the multiple parents as well as multiple rounds of inter-crossing maximising recombination and minimising population structure.
The 'NIAB Diverse' MAGIC wheat population has recently been developed from 16 founders, chosen to capture over 90% of the genetic diversity of UK adapted wheat varieties based on markers. These included both elite and historic varieties originating from several different northern European countries.
596 RILs (Recombinant Inbred Lines) have recently been genotyped at F7, and a genetic map is under development. Preliminary results from the first year of replicated yield trials indicated high power to detect QTLs in heritable traits such as flowering time and height, and further analysis on a larger number of traits including yield and yield components is ongoing. These results confirm the NIAB Diverse MAGIC population as an excellent resource for genetic dissection of complex traits.
The population is being developed as community resource. A replicated yield trial will be available for phenotyping at NIAB-Cambridge in the 2018 season, and the population is available on request. Basic phenotype data and the founder/progeny genotype data will also be available, along with the associated genetic map and QTL analysis pipeline under development at UCL. Please contact nick.fradgley@niab.com for further information.

P1001: Wheat, Barley, Oat, and related
Genome-Wide Homology Analysis Reveals New Insights into the Origin of the Wheat B Genome
Wei Zhang[1], Mingyi Zhang[1], Xianwen Zhu[1], Yaping Cao[1], Qing Sun[2], Guojia Ma[2], Shiaoman Chao[3], Changhui Yan[1], Steven S. Xu[3] and Xiwen Cai[1], (1)North Dakota State University, Fargo, ND, (2)North Dakota State University, (3)USDA-ARS, Fargo, ND
Wheat is a typical allopolyploid with three homoeologous subgenomes (A, B, and D). The ancestors of the subgenomes A and D had been identified, but not for the subgenome B. The goatgrass *Aegilops speltoides* (genome SS) has been controversially considered a candidate ancestor of the wheat B genome. However, the relationship of the Ae. speltoides S genome with the wheat B genome remains largely obscure, which has puzzled the wheat research community for nearly a century. In the present study, we performed genome-wide homology analysis to assess the B-S relationship using an integrative molecular cytogenetics and comparative genomics approach. Noticeable homology was detected between wheat chromosome 1B and Ae. speltoides chromosome 1S, but not between other chromosomes in the B and S genomes. An *Ae. speltoides*-originated segment spanning a genomic region of approximately 10.46 Mb was identified on the long arm of chromosome 1B (1BL) in all wheat species containing the B genome. The *Ae. speltoides*-originated segment on 1BL was found to co-evolve with the rest of the B genome in wheat species. Thereby, we conclude that Ae. speltoides had been involved in the origin of the wheat B genome, but should not be considered an exclusive ancestor of this genome. The wheat B genome might have a polyphyletic origin with multiple ancestors involved, including *Ae. speltoides*. These findings provide new insight into the origin and evolution of the wheat B genome, and will facilitate genome studies in wheat and its relatives.

P1002: Wheat, Barley, Oat, and related
Methods to Identify Haplotypic Variation in Wheat Cultivars and Landraces Using Multi-References and Capture Datasets
Jonathan Wright[1], Gonzalo A. Accinelli Garcia[1], Luis Yanes[1], Katie Barr[1], Laura-Jayne Gardiner[1], Philippa Borrill[2], Cristobal Uauy[2], Anthony Hall[1] and Bernardo J. Clavijo[1], (1)Earlham Institute, Norwich, United Kingdom, (2)John Innes Centre, Norwich, United Kingdom
Crop breeding programs are crucial to meet the demands of future population growth and climate change. This requires technological advances to help breeders exploit the full extent of crop genomic variation, increasing yield and disease tolerance.

Breeding programmes generate a reduction in haplotypic diversity that strengthens their key traits, but limits their overall possibilities. Crossing more genetically diverse lines, or re-introducing variation from landraces, would overcome this limitation; but identifying the right materials to cross is difficult and introducing new haplotypes is expensive and laborious.
Modern genetic and genomic tools are converging into a complete view of the genomic content of genomes. They provide new data about the available haplotypes and their distribution within target breeding programs and possible source populations, increasing the haplotypic variability available for breeders to use in a controlled and effective manner.
We present a series of methods to characterise gene content and haplotypes, combining technologies including multi-genome references, hybrid datasets (short, long and linked reads), and exome captures. We apply these methods to a variety of UK wheat cultivars and landraces, studying the haplotypic variation in different regions of the genome. We are working to identify potentially important haplotypes and show how they are exchanged between cultivars, populations and breeding programmes. We hope this will be a crucial first step towards the introduction of valuable haplotypes to help drive advances in breeding.

P1003: Wheat, Barley, Oat, and related
A Benchmarking Resource to Assess Wheat Genotyping Platforms
Mario J Caccamo, NIAB, Cambridge, United Kingdom and **Carolina Paola Sansaloni**, CIMMYT- International Maize and Wheat Improvement Centre, Texcoco, Mexico
Recent advances in both chemistry and data analysis tools have driven down the cost of DNA-based protocols resulting in a varied offer of genotyping platforms, in particular for species with complex genomes such as wheat. One of the challenges for the end users is to be able to evaluate how the different platforms compare and which ones are better suited for the different downstream applications. The aim of our project is to design a benchmarking assay to be made available to the scientific community as a resource to assess the different genotyping platforms in terms of information content, reliability of the results and resources requirements. In order to inform the construction of this resource we generated genotyping data for 384 wheat lines using three reduced-representation approaches: DArTseq, SNP-array (hybridisation-based) and exome capture. The DNA samples were extracted from material that is hosted at the seed bank at CIMMYT and includes varieties from a diversity of backgrounds (including some wild species). We will make the genotyping data publicly available via established resources (such as the CerealsDB database). We will also deploy a Galaxy interface that will provide access to the data together with a number of data analysis tools. This BBSRC Newton-fund project is a collaboration between the Seeds of Discovery initiative at CIMMYT (Mexico), and NIAB, the Earlham Institute and the James Hutton Institute in the UK.

P1004: Wheat, Barley, Oat, and related
POTAGE: A Visualisation Tool for Speeding up Gene Discovery in Wheat
Radosław Suchecki, Nathan S. Watson-Haigh and Ute Baumann, The University of Adelaide, Urrbrae, Australia
POPSEQ Ordered *Triticum aestivum* Gene Expression (POTAGE) is a web application which accelerates the process of identifying candidate genes for quantitative trait loci (QTL) in hexaploid wheat. This is achieved by leveraging several of the most commonly used data sets in wheat research. These include the Chromosome Survey Sequences, their order along the chromosomes determined by the population sequencing (POPSEQ) approach, the gene predictions and RNA-Seq expression data. POTAGE aggregates those data sets and provides an intuitive interface for biologists to explore the expression of the predicted genes and their functional annotation in a chromosomal context. The interface accelerates some of the laborious and repetitive tasks commonly undertaken in the process of identifying and prioritising genes which may underlie QTL. We illustrate the utility of POTAGE by showing how a short-list of candidate genes can quickly be identified for a QTL linked to pre-harvest sprouting - a major cause of quality and yield loss in wheat production. The candidate genes identified using POTAGE included *TaMKK3*, which was recently reported as a causal gene for seed dormancy in wheat, and a mutation in its barley ortholog has been shown to reduce pre-harvest sprouting. In addition to the public version of POTAGE, we have also developed a Docker image for quickly deploying POTAGE locally and work-flows showing how to add your own expression data sets to a local installation. This is of particular relevance to those who work with unpublished data sets or would like to deploy POTAGE on their own hardware.

P1005: Wheat, Barley, Oat, and related
GrainGenes: New Content, New Tools, New Tutorials
Victoria Carollo Blake, USDA ARS WRRC, Albany, CA, Sarah Odell, University of California, Davis, Davis, CA, Gerard R. Lazo, USDA Agricultural Research Service, WRRC, Albany, CA, Margaret Woodhouse, ISU, Ames, IA, David Hane, USDA-ARS-WRRC, Albany, CA and Taner Z. Sen, USDA -ARS / GrainGenes, Albany, CA
GrainGenes (https://graingenes.org; https://wheat.pw.usda.gov) is the USDA-ARS database for wheat, barley, oat, and rye genetics and genomics. The GrainGenes project is moving toward a genome-centric resource to accommodate the 'big data' now available for the Triticeae and Avena. In this poster, we will 1) illustrate the use of the new genome browsers on GrainGenes; 2) describe the variety-specific BLAST databases; 3) review the wealth of new content; and 4) share the collection of recently created topic-specific tutorials. Collaborations with The Triticeae Toolbox (T3), WheatIS, and Agriculture and Agri-Food Canada (AAFC) will assure that GrainGenes remains an important resource for the small grains research community. Mutual projects with our collaborators and future directions for the GrainGenes project will be discussed.

P1006: Wheat, Barley, Oat, and related
A Fast Long-Read Assembly Method for the Wheat Genome
Philipp E. Bayer[1], Chon-Kit Kenneth Chan[1], HueyTyng Lee[2], Yuxuan Yuan[1] and **David Edwards**[1], (1)University of Western Australia, Perth, Australia, (2)University of Queensland, Brisbane, Australia
Genome assembly requires the construction of continuous sequences using sequencing reads to represent the original genome sequence. Repetitive sequences within the genome are by far the largest obstacle towards obtaining gapless and accurate assemblies. The latest techniques have sought to resolve repetitive sequences using long read sequencing technologies. However, these methods demand a large amount of

computational time and high memory consumption. Here, we present a novel assembly method that reduces the runtime of current assembly methods by half, and requires around five times less memory. To demonstrate the approach, we assembled PacBio reads for a wheat genome using 300 GB of maximum RAM and 1900 hours of compute time. Our results demonstrate that this method can produce assemblies of equal or better quality when compared to current methods while using fewer resources. The success of our method to assemble this complex genome opens applications in genomic studies of other species.

P1007: Wheat, Barley, Oat, and related

Assessing the Variability of Plant Ion Transporter Genes through Time Using Modern and Ancient DNA

Oscar Estrada[1], James Breen[1], Stephen Richards[1], Geoff Fincher[2], Matthew Gilliham[1], Nana Rusishvili[3] and Alan Cooper[1], (1)The University of Adelaide, Adelaide, Australia, (2)ARC Centre of Excellence in Plant Cell Walls, Urrbrae, Australia, (3)National Museum of Georgia, Tbilisi, Georgia

Domesticated plant species such as wheat, barley, and millet have been a staple food for humans since the rise of modern agriculture about 12,000 years ago. However, cultivation and intensive breeding have also led to the loss of genetic diversity, which may threaten the continued improvement of these species. Ancient DNA opens a new path to evaluate the genetic changes occurred at different stages of domestication, by identifying potential modification, gain, and loss of critical genes involved in the plant response to the environment. Ion transporter genes are particularly influential as these loci are involved in the response of plants to abiotic stress such as salinity or drought. To assess the genetic variability of cereal species through time, we developed a plant domestication capture array that contains probes to enrich 206 nuclear genes, including ion transporter, domestication, and disease resistance related genes. Here, we present the efficiency of our array to enrich ion transporter genes in modern and ancient plant samples, including three species of wheat between 60-100 years old. From the total sequenced reads, 15% mapped to 154 target genes, representing 75% of the genes included in the array. A high proportion of sequenced reads mapped to plant ion transporter genes such as HKT and ALMT, with an average read length of 95bp and minimum coverage of 5x. Importantly, the data generated in this study can contribute valuable information for wheat breeding programs and further understanding of the evolution and domestication of this cereal.

P1008: Wheat, Barley, Oat, and related

Activation of Seminal Root Primordia during Wheat Domestication Reveals Underlying Mechanisms of Plant Resilience

Guy Golan[1], Elisha Hendel[1], Nimrod Schwartz[1] and Zvi Peleg[2], (1)Hebrew University of Jerusalem, Rehovot, Israel, (2)R. H. Smith Institute of Plant Science & Genetics in Agriculture, The Hebrew University of Jerusalem, Rehovot, Israel

Seminal roots constitute the early root system of major crops of the Poaceae family. Although variation in seminal root number was described in several crops, mechanisms through which seminal root number (SRN) are controlled and in turn contribute to environmental adaptation are poorly understood. Here, we show that SRN increased upon wheat domestication due to the activation of root primordia which are suppressed in wild wheat, a trait controlled by factors expressed in the germinating embryo. We used variation in seminal root number (SRN) between wild and domesticated wheat to investigate its bearing on water uptake and seedling resilience. The persistence of wild roots at their primordial state promoted seedling recovery from episodic water-stress through re-activation of root primordia following rehydration. In spite of their lower root number, wild seedlings transpired more than domesticated seedlings. Additionally, transpiration rate was associated with higher shoot:root ratio in wild wheat, indicating contrasting strategies of resource allocation between wild and domesticated wheat. Our findings suggest that under well-watered conditions, lower root number enables direction of resources to aboveground without limiting water uptake. Furthermore, the maintenance of roots at their primordial state and their re-activation following rehydration maybe regarded as seedling protective mechanism against episodic water-stress. The results underscore SRN as an adaptive trait that was reshaped upon domestication. Identification of factors associated with the plasticity of the SRN phenotype expands our understanding on the evolutionary dynamics of wheat and may serve to optimize root number in future breeding efforts.

P1009: Wheat, Barley, Oat, and related

Validation of Grain Yield QTL from Soft Winter Wheat Using a CIMMYT Spring Wheat Panel

Dennis Nicuh Lozada[1], Esten Mason[1], Sivakumar Sukumaran[2] and Susanne Dreisigacker[2], (1)University of Arkansas, Fayetteville, AR, (2)International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

QTL validation is an essential step in identifying genomic regions affecting variation for complex traits such as grain yield (GY) and yield components for marker-assisted breeding. We report the validation of GY-QTL from winter wheat using a population of spring wheat from CIMMYT, Mexico evaluated for GY, grain number (GNO), and thousand grain weight (TGW) across 29 international locations. The objectives of this study were to validate GY and yield component QTL previously identified from soft winter wheat using CIMMYT's wheat association mapping initiative (WAMI) panel, determine the allele combination for the validated QTL that resulted to highest GYs, and identify candidate genes associated with the validated SNP loci. KASP® assays developed for *wsnp_Ex_c361_708712* (3A), *wsnp_Ex_c13849_21698240* (4B), and *wsnp_CAP11_c3599_1741800* (6B) were associated with GY, GNO, and TGW across different BLUP and BLUE datasets in WAMI. The T-C-C allele combination, which contained favorable (positive) alleles at all three loci resulted to highest mean GY. A negative effect for the minor alleles observed in both the winter and spring panels demonstrated selection for the GY-enhancing major allele and indicated similar selection pressures in both wheat classes. Candidate gene analyses revealed diverse gene functions from repressor of RNA pol III transcription, positive regulation of ubiquitin protein ligase activity, and transcription factor identified for GY-related marker-trait associations demonstrating the complex nature of GY and yield components. Our results showed the potential of the developed assays for marker-assisted selection to improve GY-related traits in both winter and spring wheat classes.

P1010: Wheat, Barley, Oat, and related

Assessing the Consequences of Key Events of the Hexaploid Wheat Genome Evolution: Structural and Evolutive Analysis of an Ancestral Chromosomes Fusion Point and of a Region Resulting from Ancient and Recent Polyploidizations

Arnaud Bellec[1], Nadege Arnal[2], Nathalie Rodde[2], Stephane Cauet[1], Caroline Pont[3], Florent Murat[4], Jérôme Salse[4] and Helene Berges[1], (1)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (2)French Plant Genomic Resource Center - INRA Toulouse, Castanet-Tolosan, France, (3)INRA Clermont-Ferrand, France, France, (4)GDEC, INRA, Université Clermont II Blaise Pascal, Clermont-Ferrand, France
Studying the bread wheat genome's structure has been a tremendous challenge for decades. Its complexity (hexaploidy, high rate of repetitive elements) and size make wheat genome both complex to decipher and interesting to investigate as the result of multiple evolutionary forces (ancient and recent polyploidization, chromosomes fusions, deletion or sub-functionalization of homeologous genes). To study these events we first aimed at characterizing a chromosomes fusion locus (CF) and comparing the structure of the copies of a 2 Mb region carried by chromosomes 1A, 1B, 1D, 3A, 3B, 3D. Before the release of a wheat reference genome, our strategy was to identify BACs spanning the regions of interest using available resources (chromosome specific BAC libraries, physical maps, genome zipper, WGP) and to sequence BACs using PacBio technology. Despites all these resources, it remained difficult to sequence the CFs which are highly rearranged regions and the six copies of the region of interest. The release of the IWGSC reference genome followed by its annotation have changed our strategy by giving access to the whole genome sequence. Thanks to this sequence, we have redefined the orthologous relationship between hexaploid wheat genes and rice genes (as rice has conserved the structure of the common grass ancestor). We have identified various rearrangements and precisely outlined the CF regions. Following this analysis at the genome scale, we focused on the rearrangements of the CF regions of chromosomes 1A, 1B, 1D and on the fates of homeologous genes among the six copies of a region of interest.

P1011: Wheat, Barley, Oat, and related

Evaluation of the Effect of an Alien Chromosome Segment Translocated from *Aegilops sharonensis* on Recombination Frequency in Wheat

Motohiro Yoshioka, Hajime Yamada and Shuhei Nasuda, Graduate School of Agriculture, Kyoto University, Kyoto, Japan
Background: The resolution of genetic map depends on meiotic crossover frequency between homologous chromosomes. Due to a low crossover frequency in the wheat genome (*Triticum aestivum* L.), a large population is needed to obtain informative recombinants. Moreover, it has been demonstrated that artificial allotriploids in the Brassica species exhibit genome-wide elevation in recombination frequency. However, the effect of alien chromosomes on meiotic recombination still remains poorly understood in crop species. Thus, the purpose of this study is to demonstrate if the alien chromatin (segments) can increase recombination/crossover frequency in wheat.
Methods: We developed an F_2 population derived from an accession DT4B-4S^{sh}. The terminal segment of the chromosome 4S^{sh} of *Aegilops sharonensis* was translocated to chromosome 4B of Chinese Spring (CS), which was subsequently crossed with a spelta wheat (*T. spelta* L. var. duhamelianum, accession KT019-001). The F_2 population derived from normal CS and KT019-001 was used as control. Linkage maps were constructed using the same set of 494 markers scattered on A and B genome chromosomes.
Results: The complete linkage map derived from DT4B-4S^{sh} was ~11% longer than from CS. The average crossover numbers significantly increased for the entire A and B genome respectively. The observed elevation of recombination frequencies was not concentrated in a particular chromosomal region. In conclusion, our results indicated that the chromosome segment introduced from *Ae. sharonensis* increased recombination frequency globally. This suggests that small alien chromatin may affect meiotic recombination in wheat.

P1012: Wheat, Barley, Oat, and related

Deciphering Structural Variations in the Wheat Genome Using Resequencing Data

Romain De Oliveira[1], Emeric Dynomant[2], Hélène Rimbert[2], Hana Simkova[3], Jaroslav Dolezel[4], Federica Cattonaro[5], François Balfourier[2], Etienne Paux[2] and Frederic Choulet[1], (1)INRA GDEC, CLERMONT-FERRAND, France, (2)GDEC, INRA, UCA, Clermont-Ferrand, France, (3)IEB, Olomouc, Czech Republic, (4)Institute of Experimental Botany, Olomouc, Czech Republic, (5)IGA Technology Services, Udine, Italy
Structural variations (SVs) such as copy number and presence-absence variations (CNVs, PAVs) are polymorphisms that are known to be involved in the expression of phenotypes. In the absence of a reference genome sequence, their study has long been hampered in wheat. The recent advent of new wheat genomic resources has led to a paradigm shift, making possible to investigate the extent of SVs among cultivated and wild populations. Our project aims at characterizing SVs in a Triticeae diversity panel of 44 accessions from seven tetraploid and hexaploid Triticeae species. To cope with the wheat genome complexity, we developed strategies combining shotgun sequencing of sorted chromosomes 3B with bioinformatics tools and we studied SVs affecting not only genes but also transposable elements (TEs). Our results show that 14% of the genes are variable within this panel. In addition, they reveal a very high level of intra- and interspecific variability affecting TEs, contrasting with the weak polymorphism rate usually reported with SNPs. Chromosomal extremities are the regions where we see most of the variability, confirming previous hypotheses made when comparing wheat with the other grasses.

P1013: Wheat, Barley, Oat, and related

A Comprehensive Microbiome Analysis of Wheat and Its Wild Relatives

Heather N. Cantor, Gloria Broders, Kirk D. Broders and Patrick Byrne, Colorado State University, Fort Collins, CO
Microbiomes are diverse assemblages of endophytic and free-living microorganisms that can confer competitive advantages to their plant hosts such as water acquisition, nutrient mobilization, drought tolerance, salt tolerance, and disease resistance. Plant domestication and selective breeding have altered the composition of these plant-microbe interactions in several crops. It is thought that the progenitors of the A, B, and D genomes in modern hexaploid wheat manage environmental stress in their native environment by establishing symbioses with a consortium of beneficial microbes. However, these microbial communities are not well understood. The goal of this study is to better understand the core community of microbes in wild wheat relatives and how they differ from the microbiome of cultivated wheat. This study compares the bacterial and fungal taxa found in the leaves, roots, and rhizosphere of three accessions of hard winter wheat and 11 wild relatives. These plants and the agricultural soil they inhabit were sampled from a randomized complete block design with two replications, grown in well-watered and water-limited treatments in Fort Collins, Colorado. DNA was extracted and amplicon sequencing of the 16S-V4 (bacterial) and ITS2 (fungal) rRNA genes was used to describe the diversity of the microbial community associated with the root, rhizosphere and leaf of each accession.

Preliminary results indicate that while there was limited difference in microbial communities among accessions, plant compartment appears to have an important effect on structuring the microbial community across accessions.

P1014: Wheat, Barley, Oat, and related

Genome-Specific Amplicon Sequencing Strategy Provides Robust Markers across D Genome in Allohexaploid Wheat

Goro Ishikawa[1,2], Mika Saito[1], Tsuyoshi Tanaka[2], Yuichi Katayose[2], Hiroyuki Kanamori[2], Kanako Kurita[2] and Toshiki Nakamura[1], (1)Tohoku Agricultural Research Center, National Agriculture and Food Research Organization (NARO), Morioka, Iwate, Japan, (2)Institute of Crop Science, National Agriculture and Food Research Organization (NARO), Tsukuba, Ibaraki, Japan

In allohexaploid wheat, array-based genotyping platforms have been widely used for genetic analyses. However, when commercially available arrays were applied for Japanese varieties, there was a low rate of polymorphism and biased distribution of markers across the genomes, especially the D genome. These results indicate that our materials are genetically distant from those used to design the arrays. To increase genome-wide mapping markers suitable for our materials, we have established an efficient procedure to detect nucleotide polymorphisms, and a robust method for genotyping by sequencing genome and site-specific amplicons. Here we present the procedure focusing on the D genome. By sequence capture and next-generation sequencing, 12,551 polymorphisms between wheat varieties 'Hatsumochi' and 'Kitahonami' were detected across the three genomes. The flanking sequences of target polymorphisms were blasted against the International Wheat Genome Sequencing Consortium survey sequences, and three homoeologous sequences were identified. Based on the polymorphisms among the genomes, 396 D genome-specific primer pairs were designed using an in-house Java pipeline. Approximately 80% of the designed primers successfully amplified genome-specific products, indicating that they could be genotyped as easily as a diploid species. Linkage maps of recombinant inbred lines between the two varieties revealed that the newly developed markers were uniformly distributed across the D genome and greatly extended the total coverage. This result proved that the strategy described here can be useful to increase the number of markers at target sites. This work was supported by grants from the Ministry of Agriculture, Forestry and Fisheries of Japan (NGB1002, NGB1007).

P1015: Wheat, Barley, Oat, and related

Genetic Architecture of Recombination Rate and Its Effects in Allopolyploid Wheat

Katherine Jordan[1], Shichen Wang[2], Shiaoman Chao[3], Yanni Lun[1], Etienne Paux[4], Pierre Sourdille[5], Jamie Sherman[6], Alina Akhunova[7], Nancy Blake[6], Michael Pumphrey[8], Robert King[9], Andy L Phillips[9], Cristobal Uauy[10], Jorge Dubcovsky[11], Luther Talbert[6] and Eduard Akhunov[7], (1)Kansas State University, Manhattan, KS, (2)Genomics and Bioinformatics Services - TAMU, College Station, TX, (3)USDA-ARS, Fargo, ND, (4)GDEC, INRA, UCA, Clermont-Ferrand, France, (5)INRA GDEC Auvergne-Rhône-Alpes, Clermont-Ferrand, France, (6)Montana State University, Bozeman, MT, (7)Kansas State University, Department of Plant Pathology, Manhattan, KS, (8)Washington State University, Pullman, WA, (9)Rothamsted Research, Harpenden, United Kingdom, (10)John Innes Centre, Norwich, United Kingdom, (11)University of California, Davis, CA

Recombination is a natural process that shapes the landscape of natural alleles within a population. Understanding the genetic basis of how variation in recombination rate is maintained and its effects are important to manipulate the recombination process in crops in order to improve them. We dissected naturally occurring variation in recombination rate present in a spring wheat NAM population and found it is mostly defined by rare alleles with small effects that explain up to 48.3% of the variation in our population. Specifically we identified 66 regions within the genome that contribute to the observed natural variation. Our regions are enriched for meiotic functioning genes, and encompass many conserved recombination genes. Further dissection suggests that the genetic architecture of recombination is predominantly additive and controlled by trans-acting features. In addition, we observed evidence of additive genetic factors that contribute to pericentromeric crossover (CO) frequency without affecting the frequency of telomeric COs. We have also observed a negative effect of linkage drag on deleterious mutation load resulting in excess strong-effect mutations in the pericentromeric genomic regions with constrained recombination. This information suggests manipulation of wheat is possible by influencing CO distribution and frequency, thus unlocking the whole genome without the consequences for linkage drag in wheat improvement.

P1016: Wheat, Barley, Oat, and related

Optimization of Training Population for Genomic Selection

Inés Berro, Facultad de Agronomía, Universidad de la Republica, Montevideo, Uruguay, Bettina Lado, Universidad de la Republica, Montevideo, Uruguay and Lucia Gutierrez, University of Wisconsin - Madison, Madison, WI

Genomic Selection (GS) use molecular markers to predict the breeding value of individuals using a calibration set with phenotypic and genotypic information and a statistical model. The effectiveness of GS in breeding programs depends on the genetic architecture of traits, the prediction model, the number and type of markers and the structure of the training population. Our objective was to compare strategies for optimizing the training population for genomic prediction models. A total of 1535 advanced lines of wheat genotyped with 81999 SNPs was evaluated for grain yield in 35 environments. We compare three strategies to choose lines of the training population (TR) to predict a new population (TE) based on their relatedness: (1) choose lines based on the mean and median of their genetic estimated relationship matrix (G); (2) choose lines based the mean and median genetic distances calculated by the marker scores weighted by their markers effect (Gw) and (3) choose lines based on the population structure using Gw. We used two criterions for optimization: (1) the prediction error variance (PEV) and (2) the coefficient of determination (CD). Using the market effect for calculating the distance between lines was better than using the genetic relationship matrix and the median criteria was more stable than the mean. Choosing lines according to the structure of the population showed the best performance in the predicted ability. Therefore, the optimization the training population for each testing population is the best strategy to increase predict ability.

P1017: Wheat, Barley, Oat, and related

Spatio-Temporal Asymmetry of the Meiotic Pathway in Hexaploid Bread Wheat

Kim Osman, Eugenio Sanchez-Moran and Chris Franklin, University of Birmingham, Birmingham, United Kingdom

During meiosis homologous recombination (HR) generates genetic variation and provides the physical links (crossovers - COs) necessary for accurate segregation of chromosomes. In most eukaryotes the distribution of COs along chromosomes is non-random due to the influence of multiple levels of control which ensure each chromosome pair receives at least one CO and which discourage additional COs forming in adjacent chromosomal regions. Further complexity is evident in the tendency of chiasmata (the cytological manifestation of COs) to form in favourable regions of the chromosome. In some species this has led to the restriction of COs/chiasmata to particular chromosomal locations. In hexaploid wheat and other cereals the predominantly distal location of COs creates a problem of linkage-drag in the recombinationally 'cold' centromere/proximal and interstitial regions where agronomically important traits cannot be readily separated from undesirable ones.
As partners in a collaborative project involving five UK research groups and two wheat-breeding companies, our aim is to understand the factors influencing CO formation in hexaploid wheat in order to manipulate the process and unlock genetic diversity for crop improvement. Building on research in Arabidopsis meiosis, we are employing molecular cytogenetic techniques to perform a detailed analysis of key stages in the recombination pathway during the progression of prophase I. Here we present data showing that early recombination events in Cadenza are spatio-temporally asynchronous, initiating in the distal chromosomal regions and later spreading throughout the chromatin. This pattern reflects the distribution of euchromatin within the nucleus as revealed by immunolocalisation of various histone modifications.

P1018: Wheat, Barley, Oat, and related
Genetic Basis of the Short Life Cycle of 'Apogee' Wheat
Carol Powers[1], Rungravee Boontung[2], P. Stephen Baenziger[2] and Liuling Yan[1], (1)Oklahoma State University, Stillwater, OK, (2)Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE
'Apogee' is a wheat cultivar that was developed for utilization of the NASA-ALSS food system and has the shortest life cycle in wheat in the world, with flowering only 25 days after planting under long day conditions and constant warm temperature without vernalization. This growth habit can be utilized to accelerate breeding cycles. It is intriguing to unravel the genetic mystery of this agronomic characteristic. In this study, Apogee was crossed with a strong winter wheat cultivar 'Overland', and over 800 F2 plants were generated and tested in a greenhouse under temperature and photoperiod controlled conditions. Apogee was found to have *vrn-A1a* and *vrn-D3a* that are the same alleles as observed in the winter wheat cultivar 'Jagger', *Vrn-B1* that has a deletion in intron one, and *PPD-D1b* that is insensitive to photoperiod. The super-short life cycle of 'Apogee' wheat resulted from pyramiding of the early alleles for the four flowering time genes, whose effects are *vrn-A1>VRN-B1>vrn-D3>PPD-D1*. The dominant *vrn-D3a* alone was not sufficient to induce the transition from vegetative to reproductive development in winter plants without vernalization, but did accelerate heading in those plants that have been induced by *vrn-A1a* or *Vrn-B1*. This study greatly advanced the molecular understanding of the multiple flowering genes under different genetic backgrounds and provided useful molecular tools that can be used to accelerate winter wheat breeding schemes.

P1019: Wheat, Barley, Oat, and related
A TRIM Insertion led to a Gene Resurrection Event that causes Male Sterility in Wheat
Jizeng Jia, Chinese Academy of Agricultural Sciences, Beijing, China
The male sterile *ms2* mutant has been known for 40 years[1] and has become extremely important in the commercial production of wheat. However, the gene responsible for this phenotype has remained unknown. We here report the map-based-cloning of the *Ms2* gene. The *Ms2* locus is remarkable in several ways that have implications in basic biology. Beyond having no functional annotation and clearly having undergone pseudogenization, we found that the *Ms2* allele in the *ms2* mutant acquired a terminal-repeat retrotransposon in miniature (TRIM) element in its promoter. This TRIM element is responsible for the anther-specific *Ms2* activation that confers male sterility. The identification of *Ms2* not only unravels the genetic basis of a historically-important breeding gene, but also illustrates pseudogenization at the population level and shows that resurrection of an unfixed pseudogene in the population can contribute to genetic novelty and phenotypic plasticity.

P1020: Wheat, Barley, Oat, and related
Exploring Epigenomic Diversity in Polyploid Wheat
Laura-Jayne Gardiner[1], Ryan Joynson[1], Jimmy Omony[2], Rachel Rushholme-Pilcher[1], Lisa Olohan[3], Daniel Lang[4], Caihong Bai[5], Malcolm Hawkesford[5], David Salt[6], Manuel Spannagl[7], Klaus F. X. Mayer[8], John Kenny[3], Michael Bevan[9], Neil Hall[1] and Anthony Hall[1], (1)Earlham Institute, Norwich, United Kingdom, (2)HelmholtzZentrum Munchen, Munich, Germany, (3)University of Liverpool, Liverpool, United Kingdom, (4)Helmholtz Center Munich, Neuherberg, Germany, (5)Rothamsted Research, Harpenden, United Kingdom, (6)University of Aberdeen, Aberdeen, United Kingdom, (7)PGSB - Plant Genome and Systems Biology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany, (8)Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany, (9)John Innes Centre, Norwich, United Kingdom
Wheat has been domesticated into a large number of agricultural environments, a key question is what drives the ability for crops to rapidly adapt. To address this question, we survey genotype and DNA methylation across the core Watkins bread wheat landrace collection that is representative of global wheat genetic diversity. We identify independent variation in methylation, genotype and transposon copy number. These three sources of variation are likely to be driving phenotypic differences across this diverse wheat collection. Methylation and transposon diversity could therefore be used alongside single nucleotide polymorphism (SNP) based markers for breeding.

P1021: Wheat, Barley, Oat, and related
The Interplay Among Subgenomes Shapes Genomic Variations and Transcriptomic Changes during Wheat Hexaploidization Events
Mingming Xin, CHINA AGRICULTURAL UNIVERSITY, BEIJING, China
Genomic variations and transcriptomic changes extensively occur in newly formed polyploids to reconcile immediate challenges caused by divergent subgenomes in one nucleus. To comprehensively investigate sequence elimination and expression alteration in wheat hexaploidization, here we performed whole exome capture experiments coupled with high throughput sequencing analysis by exploiting three

sets of newly synthesized wheat species. We observed that the whole wheat chromosomes was subjected to extensive genomic elimination partially regulated by sequence homology in response to hexaploidization. But homeologous subgenomes exhibited distinct features that DNA sequences were preferentially eliminated on DD genome compared with AA and BB genome. In addition, a higher proportion of eliminated sequences occurred in exonic regions than in intergenic regions on DD genome, whereas a significant enrichment was observed in the repeat-rich intergenic regions for AA and BB genome, exhibiting a contrast distribution pattern compared to the gene density. Furthermore, we detected 488 overlapped genes with sequence elimination on DD genome but few on the other two genomes across three nascent hexaploid wheats. Interestingly, GO enrichment analysis showed genes with sequence elimination were enriched in distinct functional pathways between subgenomes. Transcriptome analysis indicates polyploidization enhanced gene expression differentiation between root and leaf and led to rapid and extensive gene expression changes in synthetic hexaploid wheat. AA and BB genome exhibited synergistic expression profiling which was distinct from DD genome, and interestingly, expression bias was observed for a proportion of homeologs in synthetic hexaploid wheat. Strikingly, only 3.3-23.6% genes with sequence elimination exhibited expression changes in synthetic hexaploid wheat compared with their respective progenitors, indicating genomic variation is not the major cause resulting in the transcirptomic changes during wheat polyploidization events, whereas epigenetic modifications might play an important role in regulating expression profiling alterations.

P1022: Wheat, Barley, Oat, and related

Analysis of QTL Associated with Yield and Yield Components in TAM 111 and TAM 112 and their Interactions with Environments

Yan Yang[1,2], Smit Dhakal[2], Jackie C Rudd[1], Qingwu Xue[1], Amir M. Ibrahim[2], Kirk Jessup[1], Lisa Garza[1], Ravindra Devkota[1], Jason Baker[1], Shannon Baker[1], Shichen Wang[3], Richard Metz[3], Charles D. Johnson[3] and Shuyu Liu[1], (1)Texas A&M AgriLife Research, Amarillo, TX, (2)Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, (3)Genomics and Bioinformatics Services - TAMU, College Station, TX

A population of 124 recombinant inbred lines (RILs) was developed from two popular hard red winter wheat (HHRW) cultivars TAM 111 and TAM 112 to identify and characterize QTL for yield and yield components in wheat (*Triticum aestivum* L.). A set of 9928 markers derived from the wheat 90K iSelect array and genotype-by-sequence (GBS) was used to genotype the population and constructed high-density genetic map for QTL analyses. Data for yield and yield components was obtained from 8 environments with various year, location and irrigation. QTL analysis was performed using the software MapQTL, GenStat, and QTLNetwork based on eight individual environments and three mega-environments. Four unique and consistent QTL regions with pleiotropic effects were identified after comparing different models, which were distributed on chromosome fragment 1D2, 2D1, 4D, and 7D1. The QTL on chromosome 1D and 2D were associated with thousand kernel weight (TKW). The 4D QTL was associated with grain yield (GY) and TKW. The QTL detected on 7D chromosome linked with GY, TKW, kernel per spike (KPS), and harvest index (HI). The sequence of one SNP linked to QTL on 7D, was identical to wheat mRNA for starch synthase I-1 (*wSsI-1* gene) based on BALST analyses. This protein is involved in the pathway starch biosynthesis, which is part of Glycan biosynthesis. The high-saturated genetic map and the information on the candidate genes annotation could be useful in map based cloning and marker development for marker-assisted breeding for higher yield components.

P1023: Wheat, Barley, Oat, and related

A Genome Wide Association Study for Yield Traits in Soft Red Winter Wheat

Dylan Lee Larkin, University of Arkansas, Fayetteville, AR

Wheat (*Triticum aestivum* L.) is a widely produced grain crop, significantly contributing to global food security. As the global population continues to grow, so will the demand for food. In order to meet such demands, breeders must work to increase wheat yield potential. Wheat yield can be impacted by multiple quantitative traits which rely on several quantitative trait loci (QTL). A genome wide association study (GWAS) was conducted on 360 inbred soft red winter wheat genotypes adapted to the southern United States in an association mapping panel (AMP) in order to identify novel QTL associated with wheat yield traits, including yield, test weight, heading date, maturity date, and plant height. The AMP was grown over eight location-years between 2013 and 2017 in randomized complete block and augmented designs. Each location-year was evaluated for the five aforementioned traits impacting yield. Best linear unbiased estimates for the five traits were obtained from a spatial linear mixed model for each location-year and combined to obtain best linear unbiased predictions from SAS 9.4 software. Genotype-by-sequencing (GBS) identified 71,428 high quality single nucleotide polymorphisms (SNP) markers across all 21 wheat chromosomes. Marker-trait associations will be determined using the FarmCPU function in R software. Data analysis for the five phenotypic traits are still in progress. Marker-trait associations will also be performed in the near future, resulting in potential SNP that can be implemented by the University of Arkansas wheat breeding and genetics program through marker assisted selection or genomic selection in order to improve wheat yield.

P1024: Wheat, Barley, Oat, and related

Fine Mapping of D-Genome Yield QTLs in Hexaploid Wheat

Jonathan D Turkus[1], Eric Olson[1] and Sunish Sehgal[2], (1)Michigan State University, East Lansing, MI, (2)South Dakota State University, Brookings, SD

Increasing crop yields is an ever more crucial endeavor as the global human population continues its near exponential growth. One strategy to meeting future food demands is identifying and incorporating yield-enhancing genes into elite crop lines. Wheat (*Triticum aestivum*, 2n=6x=42 AABBDD), which supplies a fifth of humans' calories worldwide, can reap the benefits of the genetic variation from the D-genome progenitor, *Aegilops tauschii* (2n=2x=14 DD), which may supply yield-boosting alleles. A nested association mapping population of the D-genome (DNAM) was created from direct hybridization of the hard white winter variety, KS05HW14, and seven *Ae. tauschii* accessions and backcrossing the F1 progeny twice to KS05HW14 to regain euploidy. BC_2F_4 derived lines from the DNAM population were phenotyped for grain yield in Manhattan, KS; Hays, KS; and Richville, MI in 2015 and 2016, and Marianna, AR; Champaign, IL; Brookings, SD; and Pullman, WA in 2016. A genome-wide association analysis identified QTL conferring higher grain yield. Large-effect QTL identified on chromosomes 2DS and 6DL were contributed by the recurrent parent. QTL on chromosomes 2DL and 7DS were derived from the Ae. *tauschii*

accessions, TA1615 and TA1718. KASP markers designed for significant SNPs on the QTL identified the following segregating regions: 25.4Mb to 29.5Mb on 2DS, 430.4Mb to 575.9Mb on 2DL, 463.5Mb to 473.3Mb on 6DL, and 517.0Kb to 12.5Mb on 7DS. These data were used to create heterozygous inbred families that will be used in QTL fine-mapping and identifying yield-enhancing genes.

P1025: Wheat, Barley, Oat, and related

Breeding Strategies using Genomic Selection Increase Genetic Gain in Wheat Breeding Programs

Biructawit B. Tessema[1], Huiming Liu[1], Anders Christian Sørensen[1], Jeppe R. Andersen[2] and Just Jensen[1], (1)Molecular Biology and Genetics, Aarhus University, Tjele, Denmark, (2)Nordic Seed, Galten, Denmark

The traditional wheat breeding programs have been running for several years yet the genetic gain has been very limited. However, the use of genomic information for a selection criterion can increase genetic gain. This study was set to see how much genetic gain can be increased by implementing genomic selection on traditional wheat breeding program. In addition, we investigated the effect of genetic correlation between different traits on genetic gain. A series of wheat breeding programs that run simultaneously for 30 years was simulated using stochastic simulation, meaning each year a new breeding program starts with a cross of 60 parental lines followed by six generations of selfing. Selection was performed on three different generations. At F2, phenotypic selection was performed on breeder's visual preference. At F5 and F6, either phenotype or Genomic Estimated Breeding Value (GEBV) was used to select on yield. Yield at F5 and F6 was considered as different traits because they differ in plot size, population density, and number of plot replications. Plot heritability of these traits were 0.1, 0.2, and 0.3 while the economic values were 0, 0, and 1. In addition, we simulated different levels (0.3, 0.5, 0.7, 0.9) of genetic correlation between F2 and F5 as well as between F5 and F6. The varied selection criterion and varied genetic correlations make a total of 16 scenarios. GEBV as a selection criterion significantly increased genetic gain by 10% compared to phenotype. Besides, the genetic gain was higher with the higher genetic correlation between traits.

P1026: Wheat, Barley, Oat, and related

Genetic Analysis of Grain Shape, Grain Weight, Test Weight, Milling Yield, and Plant Height in a Spring Wheat Cross

Curt McCartney, Agriculture and Agri-Food Canada, Morden, MB, Canada, Adrian Cabral, National Research Council of Canada, Sakatoon, SK, Canada, Mark Jordan, Agriculture & Agri-Food Canada, Morden, MB, Canada, Gary Larson, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada, Daryl J. Somers, Vineland Research and Innovation Centre, Vineland station, ON, Canada and Gavin Humphreys, Agriculture and Agri-Food Canada, Ottawa, ON, Canada

Wheat grain shape affects traits under selection by breeders (e.g. grain weight, test weight, and milling yield). A doubled haploid (DH) population of the cross RL4452/'AC Domain' was used to study the genetic basis of seed shape. Quantitative trait loci (QTL) analyses were conducted on a total of 18 traits: 14 grain shape traits, plant height, 1000-grain weight, test weight, and milling yield. Grain samples were harvested from trials grown at Glenlea, Brandon and Morden in Manitoba, Canada, between 1999 and 2004. Kernel shape was studied through digital image analysis with an Acurum® grain analyzer. Plant height, grain weight, test weight, flour yield, and grain shape were correlated with each other and QTL analysis revealed that QTL for these traits often mapped to the same genetic locations. The most significant QTL for grain shape traits were located on chromosomes 4B and 4D coincident with QTL for plant height. The most significant QTL for plant height, grain weight, and test weight mapped to the *Rht-D1* locus on chromosome 4D. *Rht-D1b* decreased plant height, grain weight, test weight, and kernel width relative to the *Rht-D1a* allele. A narrow genetic interval on chromosome 4B contained significant QTL for grain shape, grain weight, and plant height. The 'AC Domain' allele reduced plant height, grain weight, kernel length and width traits, but had no detectable effect on test weight. The cumulative data indicated that this variation was inconsistent with segregation at *Rht-B1*. Numerous additional QTL were also identified that control these traits in this population.

P1027: Wheat, Barley, Oat, and related

Development and Validation of a Single Nucleotide Polymorphic Marker for the Yield Component Kernel Weight in Wheat

Santosh Kumar[1], Sudhakar Pandurangan[2], Jacob Toth[2], Andrew Burt[2] and Jennifer Mitchell Fetch[2], (1)Agriculture and Agri-Food Canada, Brandon, MB, Canada, (2)Agriculture and Agri-Food Canada, Canada

Canada is a major producer and exporter of hard red spring (HRS) wheat. The HRS wheat is well known for its excellent milling and baking quality. With high protein requirement in the HRS class of wheat, it lags in yield compared to other classes of wheat. Yield is of utmost importance to producers and is one of the primary focuses of wheat breeding programs. An important yield component is the thousand kernel weight (TKW) and is highly heritable. Selection for high TKW in early generations of wheat breeding is effective, but is difficult to access phenotypically due to limited seed availability. Marker-assisted selection (MAS) using single nucleotide polymorphisms (SNPs) and insertion-deletion (indels) mutations will allow selection in early generations breeding lines based on genotype. Kompetitive Allele Specific PCR (KASP) assays based on SNPs and indels are high-throughput, easy to use, and requiring limited amounts of DNA. Trehalose-6-phosphate (T6P) is a regulator of starch accumulation, the most important contributor of TKW. Trehalose 6-phosphate phosphatase (T6PP) activity can affect T6P levels, and a wheat *T6PP* gene has been cloned. This gene was found to be polymorphic in Chinese accessions of wheat using a cleaved amplified polymorphic sequence assay and linked to TKW. Here we developed a much simpler and high-throughput KASP assay and showed that this gene is polymorphic in Canadian wheat germplasm. This should allow selection for TKW in early generations of wheat breeding in Canada.

P1028: Wheat, Barley, Oat, and related

Characterization, Validation, and Deployment of Chromosome 6BL and 7AL QTLs for Grain Yield Components in Hard Winter Wheat

Andrew M. Katz, Patrick Byrne, Scott Haley and Stephen Pearce, Colorado State University, Fort Collins, CO

The United Nations has estimated that food production will need to double by 2050 to adequately feed a global population of 9 billion people. Improvements in wheat yields, which account for 30% of coarse grain production, will be essential to meet this goal. Yield is a complex trait

due to a multitude of influential factors. To address this complexity we have identified individual yield components that are less complex and contribute to overall yield. A GWAS of a hard winter wheat association-mapping panel identified QTLs on the 7AL chromosome arm for spikelet number and the 6BL chromosome arm for kernel width. The Great Plains winter wheat cultivar Platte and experimental line CO940610 were identified as polymorphic in the 7AL and 6BL regions. A population of recombinant inbred lines was generated from the two parents and used to validate the 7AL and 6BL QTLs' effects. Individual SNPs have been identified which will be used to introgress spikelet number and kernel width QTL into Colorado advanced lines and high biomass lines from the International Maize and Wheat Improvement Center (CIMMYT). Exome sequencing data generated from the parental lines will enable high-resolution mapping of the causative genetic variant underlying these QTL. The employment of novel genomic tools and resources enable unprecedented opportunities to identify allelic variation underlying individual yield components in wheat. This will ultimately aid in the development of higher yielding wheat varieties.

P1029: Wheat, Barley, Oat, and related

Resource Optimization with Multi-Trait Genomic Prediction for Bread Wheat Quality

Bettina Lado[1], Daniel Vazquez[2], Paula Silva[3,4], Martin Quincke[2] and Lucia Gutierrez[1,5], (1)Universidad de la Republica, Montevideo, Uruguay, (2)Instituto Nacional de Investigación Agropecuaria, Colonia, Uruguay, (3)Department of Plant Pathology, Kansas State University, Manhattan, KS, (4)Kansas State University, Manhattan, KS, (5)University of Wisconsin - Madison, Madison, WI

Multi-trait genomic prediction models are a useful strategy to predict traits that otherwise are challenging due to labor intensity, difficulty, and cost. This is particularly important in the context of resource allocation in plant breeding programs. However, is not well known the amount of phenotyping that could be replaced by including phenotypic information on correlated traits. The objective of this work was to compare the predictive ability of multi-trait models, 1) by using different training population sizes for different quality traits, and 2) by testing different proportions of lines with phenotypic information for correlated traits. A group of 495 wheat lines were genotyped using genotyping by sequencing and phenotyped for eight bread quality traits. Cross-validation was used to evaluate the predictive ability of different multi-trait models using 10 to 80% of lines as training population and 50, 75 or 100% information on correlated traits. The results showed that predictive ability for all traits did not change when using more than 30% of lines as training population and 100% of the information on correlated traits. Moreover, the predictive ability of multi-trait models decreased when information on correlated traits was reduced to 50%. Overall, our results indicate that inclusion of information on correlated traits in training and testing wheat lines is a useful approach to replace phenotyping of expensive traits, allowing to reduce costs and better allocate resources in breeding programs.

P1030: Wheat, Barley, Oat, and related

Effect of Glutenin Genes and Glutenin Gene by Environmental Interaction on Quality in Spring Wheat

Emily Delorean[1], Carlos Guzmán[2] and Jesse Poland[1], (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Wheat quality, comprised of milling yield, dough and baking quality, is a critical objective in the spring wheat breeding program at the International Maize and Wheat Improvement Center (CIMMYT). The glutenin genes encode the proteins that are part of the gluten matrix that gives rise to strength, extensibility and elasticity for which wheat dough is famous and are known to be an important determinate in processing and end-use quality. Previous studies have found significant genotype, environment and genotype by environment effects on grain quality traits. Recently, these same trends were also found in a population of 56 hard spring wheat lines grown in 6 environments representing 2 levels each of irrigation, drought or heat stress. It was hypothesized that glutenin genes were underlying the significant genotype effects and glutenin alleles may interact with environments. The effects of glutenin alleles as well as the environment by glutenin effects on quality were tested with a mixed linear model on this population. Glutenin loci were found to have a significant association with most grain quality phenotypes, with the high molecular weight glutenins (*Glu-A1*, *Glu-B1*, and *Glu-D1*) having larger effects. Additionally, the glutenin by environment interaction was significant for some of the glutenin loci. The results of this study confirm that glutenin alleles do underlie some of the genotype effect on quality traits in wheat and different glutenin alleles are performing different in contrasting environments. This information can help breeders at CIMMYT to target wheat quality profiles of lines to specific environments.

P1031: Wheat, Barley, Oat, and related

Reduced Height Semi-Dwarf Alleles Significantly Impact Wheat End Use Quality

Emma Jobson[1], John M. Martin[2], Luther Talbert[1], Tavin Schneider[1] and Mike Giroux[1], (1)Montana State University, Bozeman, MT, (2)Montana State Univ, Bozeman, MT

The genes responsible for dramatic yield increases during the 1950s and 1960s are mutant forms of the *Reduced height* (*Rht*) genes. Since then, two semi-dwarfing alleles of *Rht*, *Rht-B1b* and *Rht-D1b* have been widely incorporated into modern wheat cultivars. Despite being some of the most widely utilized yield increasing genes, few studies have examined their effect on wheat end use quality. For this study we compared near isogenic lines created in a standard height spring wheat variety (Fortuna) varying for the presence of the gibberellin insensitive mutations *Rht-B1b, Rht-D1b,* or the gibberellin sensitive dwarfing gene *Rht-8*. Our agronomic results agreed with previous findings and we observed a 25% height reduction and 13% yield increase in *Rht-B1b/Rht-D1b* compared to the tall isoline. Grain protein was decreased (from 15.4 to ~13.6%) as was kernel weight (15%) in the *Rht-B1b/Rht-D1b* isolines. We also saw a slight decrease in the loaf volume in *Rht-B1b/Rht-D1b* compared to the tall line. However, we observed statistically significant increases in flour yield (2%), falling number (5%), mixing tolerance (56%), and baking mix time (33%). The fact that flour yield was increased is unexpected since *Rht-B1b/Rht-D1b* also decrease seed size. For almost all parameters, *Rht-8* was intermediate between the tall line and *Rht-B1b/Rht-D1b*. These findings indicate that although *Rht-B1b/Rht-D1b* decrease seed size and protein content, they positively impact wheat milling yield and do not negatively impact dough strength.

P1032: Wheat, Barley, Oat, and related

Genomic Selection for End-Use Quality Traits in Soft White Wheat (Triticum aestivum L.)

Jayfred Gaham Godoy[1], Meng Huang[1], Zhiwu Zhang[2] and **Arron H. Carter**[1], (1)Washington State University, Pullman, WA, (2)Dept. of Crop and Soil Science, Washington State University, Pullman, WA
End-use quality traits in soft white wheat are complex traits that are controlled by multiple genetic factors with minor effects. A previous genome-wide association study (GWAS) identified 105 SNP markers for end-use quality traits but these markers only explained 5 – 30% of the phenotypic variation leaving a larger portion of unaccounted heritability. Genomic selection (GS) is a breeding method to predict breeding values using genome-wide markers. GS can simultaneously model all additive genetic variance that is unaccounted for in GWAS. We assessed the application of GS for 21 end-use quality traits using a panel of 469 elite soft white winter wheat from Pacific Northwest breeding programs that were genotyped with 15,229 SNP markers. Genomic prediction using single and multi-trait models were evaluated using the R packages rrBLUP and PHENIX, respectively. Single trait prediction estimates were calculated using the gBLUP model. The multi-trait model used genetic information from the kinship matrix and trait correlation to estimate genomic estimated breeding values (GEBVs). Prediction accuracies following a 10-fold cross validation were 30 – 87% for the single trait model and 69 – 99% for the multi-trait model. Prediction accuracies were significantly higher (up to a 100% increase) in the multi-trait model especially for low heritability traits. Our results suggest that genomic selection can be an efficient tool to develop soft white wheat with superior end-use quality traits. We are currently validating the multi-trait GS model to predict end-use quality performance in different breeding populations (e.g. F5 single plots and double haploids) using genotype-by-sequencing data.

P1033: Wheat, Barley, Oat, and related

Analysis of Wheat Internode Time Course RNAseq Data to Identify Genes Related to Nutrient Remobilisation Depending on Nitrogen Availability

Asier Gonzalez-Uriarte, David J Hughes, Peter Buchner and Keywan Hassani-Pak, Rothamsted Research, Harpenden, United Kingdom
The massive and polyploid genome of wheat has make it challenging for researchers to leverage RNAseq to identify candidate genes underlying complex phenotypes. Most studies in wheat have focused on high-level characteristics of gene expression because the available references were incomplete and fragmented and over 100,000 transcribed genes are assayed simultaneously. In other organisms, well designed and carefully analysed RNAseq experiments have contributed to gain insights into mechanisms controlling complex traits and we intend to follow a similar approach in wheat genomics now that near-complete references are available.
Here we describe the bioinformatics analysis of a time course RNAseq experiment in which internode samples from hexaploid wheat (*T. aestivum*) plants treated with two nitrogen levels have been collected at 8 time points between anthesis and senescence. The study aims to elucidate the molecular mechanisms involved in nutrient remobilisation from senescing organs to developing tissues and the role of nitrogen availability in senescence. The chloropyll level has been used as a marker of senescence and the DESeq2 Bioconductor package has been used to identify genes that switch their state at its onset.
We demonstrate that it is possible to reduce the number of candidate genes from thousands to dozens using statistical methods to identify genes that have an expression profile of interest. In this work, 9 genes have been found to be upregulated and 8 downregulated in senescence and their functions seem to be related to processes like chlorophyll binding, protein breakdown and cell signalling.

P1034: Wheat, Barley, Oat, and related

Nitrogen Use Efficiency Is Regulated By Interacting Proteins Relevant to Development in Wheat

Lei Lei, Carol Powers, Brett Carver, Liuling Yan and Hailin Zhang, Oklahoma State University, Stillwater, OK
Nitrogen (N) is the most important nutrient for plant development and growth, and soil is often supplemented with N fertilizer to ensure successful seed production and high grain yield for non-N-fixing food crops such as wheat (*Triticum aestivum* L.). Only 30–35% of added N fertilizers are taken up and used by wheat plants in the year of application, and the remaining 65–70% (assuming fertilizer–soil equilibrium) is lost. Developing varieties of wheat that require less N input yet maintain the same or higher grain yield is an economically and environmentally sustainable goal in international agriculture. In this study, a major quantitative trait locus (QTL) for N-related agronomic traits was cloned from wheat. The vernalization gene *TaVRN-A1* was tightly linked with the gene at the QTL. Due to the Ala^{180}/Val^{180} substitution, *Ta*VRN-A1a and *Ta*VRN-A1b proteins had differential interactions with *Ta*ANR1 protein, which is encoded by a wheat orthologue of *Arabidopsis nitrate regulated 1* (*ANR1*). A natural mutant of *TaANR1* was found which is missing exon 6 in its mRNA, which had genetic effect on wheat development and growth. The transcripts of both *TaVRN-A1* and *TaANR1* were down-regulated by N. Genetically incorporating favorable alleles from *TaVRN-A1*, *TaANR1*, and *TaHOX1* increased grain yield from 9.83% to 11.58% in a winter wheat population tested in the field.

P1035: Wheat, Barley, Oat, and related

Indo-UK Research Collaboration to Improve Nitrogen Use Efficiency in Wheat

Soma Marla, ICAR.NBPGR, New Delhi, India
Nitrogen is the major agronomic input that determines the performance and productivity of wheat crop in both India and UK. With nitrogen being the major production cost for farmers, it has a huge environmental footprint, in terms of pollution of ground waters and generation of greenhouse gases. To minimize the use of applied Nitrogen fertilizers a cross-Institute pre-breeding programme (INEW virtual centre) under India-UK partnership is being executed to identify sources of traits and developing markers for use in academic research and transfer to commercial breeding programmes and responsibility for delivering improved wheat varieties to Indian farmers. The Virtual Joint Centre is bringing together major wheat researchers from ICAR.IARI in New Delhi, Indian Institute of Wheat and Barley Research, Karnal, Bourlag Institute for South Asia, Punjab, National Bureau of Plant Genetic resources, New Delhi and National Research Centre for Plant Biotechnology, New Delhi,Punjab Agricultural University. UK partner institutions are Rothamsted research, University of Nottingham, University of Bristol, John Inns Centre and National Institute of Agricultural Botany.
The core of the project is precision field trials being conducted in India and UK using germplasm from both countries, in which the fate of nitrogen in the plant will be followed from root uptake to seed maturity, at limiting and adequate levels of fertilisation. Its impact on grain yield and grain quality in these lines will be studied in detail using the platform technologies (Fig.1) providing information on the relationship between performance and phenology. Major mandate includes integrated study of the genetic, biochemical and molecular basis for improved N

use efficiency from mechanisms of nitrogen uptake to partitioning in the grain and effects on processing quality. Candidate genes that control key processes limiting N use efficiency will be identified. The study will be supported by genotyping of germplasm and identification of key genes, enzymes involved, their variation employing high density SNP arrays and transcriptome analysis. Molecular markers developed for key traits will be transferred to wheat breeders in UK and India.

P1036: Wheat, Barley, Oat, and related

Cytogenetic Analysis of Karyotypically Unstable Perennial Wheat Amphiploid Breeding Lines

Matthew Arterburn, Washburn University, Topeka, KS, Colin Curwen-McAdams, Washington State University, Burlington, WA and Stephen Jones, Washburn State University, Burlington, WA

Perennial wheat breeding lines are amphiploids generated by wide hybridization of annual hexaploid bread wheat (*Triticum aestivum*, 2n = 6x = 42, AABBDD) with perennial wheatgrass species such as tall wheatgrass (*Thinopyrum elongatum*, 2n = 14, E^eE^e) and intermediate wheatgrass (*Thinopyrum intermedium*, 2n = 6x = 42, $E^eE^eE^eE^eStSt$) and doubling chromosome content with colchicine. This process mimics polyploidization events that have occurred throughout evolutionary history in the plant tribe Triticeae. Such lines demonstrate a perennial habit and are useful in sustainable agriculture systems. Post-hybridization generations of these lines exhibit considerable karyotypic instability, and chromosome loss results. Similarly, when perennial wheat amphiploids are crossed to other wheat cultivars as part of breeding strategies, subsequent generations experience considerable chromosome number variation. To characterize chromosome behavior in early and late generations after amphiploid crosses, and to distinguish perennial wheat breeding lines with promising agronomic characteristics, we performed cytogenetic analysis on 38 lines selected from multiple generations (F1 through F6) after initial crosses involving wheat and various wild perennial wheatgrass species from the *Thinopyrum* genus. We used genomic in situ hybridization (GISH) to identify the genome origins of the chromosomes present. Among the various lines analyzed, chromosome numbers ranged from 43 to 70, with GISH analysis indicating great variation among the alien chromosomes. Chromosomal aberrations were evident in a number of lines, including telosomes and translocations between wheat and the alien chromosomes. We also employed molecular marker analysis, using cleaved amplified polymorphic sequence (CAPS) markers, to help specifically identify alien chromosomes present.

P1037: Wheat, Barley, Oat, and related

Exploring Genetic Diversity in Bread Wheat Using Nested Association Mapping

Melissa Garcia[1], Robert Pontre[2], James Edwards[3], Kerrie Forrest[4], Debbie Wong[5], Matthew Hayden[6], Tristan Coram[3], Dan Mullan[7] and Delphine Fleury[8], (1)School of Agriculture, Food and Wine, University of Adelaide, Urrbrae, Australia, (2)University of Adelaide, School of Agriculture, Food and Wine, Urrbrae, Australia, (3)Australian Grain Technologies, Adelaide, Australia, (4)Department of Economic Development, Jobs, Transport and Resources, Melbourne, Australia, (5)DEDJTR, Biosciences Research, AgriBio, Melbourne, Australia, (6)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia, (7)Intergrain Pty Ltd, Perth, Australia, (8)University of Adelaide, Adelaide, Australia

The rate of genetic gain in breeding programs can be increased by extending the amount of variation available for selection using land races and exotic germplasm. However, exotic germplasm carries a range of undesirable traits that limits their suitability for modern agriculture. Backcrossing to locally adapted varieties and pre-selection for traits is therefore required to ensure meaningful data are generated in field trials. Multi-parental schemes such as Nested Association Mapping (NAM) populations improve the use of exotic germplasm as a resource for the discovery of novel traits and QTL/genes. NAM combines the power of linkage analysis and the precision of association mapping. When jointly analysed, NAM populations can provide higher power to detect QTL than any of the constituent biparental families separately.

We selected 75 highly diverse hexaploid spring-type wheat accessions from regions of the world that are affected by heat and drought stress. These accessions were crossed with two Australian elite varieties as founder parents, and BC_1F_6 populations were generated.

Twenty individuals from each of 28 NAM sub-populations were genotyped at BC_1F_4 using a targeted genotype by sequencing assay. These lines are being grown in a completely randomised field trial with two replications. Plots were phenotyped for NDVI, relative maturity and presence of awns. Plant height, yield, thousand grain weight and harvest index will be obtained at harvest.

Genome wide association analysis is underway. Selected populations which maximize diversity and power will be phenotyped in multiple field trials across Australia next year.

P1038: Wheat, Barley, Oat, and related

Vitamin A Biofortification of Wheat Grains using a TILLING Mutant-Based Approach

Shu Yu[1], Xiaoqiong Qin[1], Jorge Dubcovsky[2] and **Li Tian**[1], (1)University of California, Davis, Davis, CA, (2)University of California, Davis, CA

Vitamin A deficiency (VAD) has been widely recognized as a major public health problem in many parts of the world. While wheat provides about 20% of dietary calories and proteins worldwide, wheat grains (particularly endosperm/flour) are generally low in vitamins (e.g. provitamin A) and minerals. We are breeding for increased accumulation of beta-carotene, a provitamin A molecule, in tetraploid wheat grain endosperm using induced mutations. Our spatial gene expression analysis indicated that specific carotenoid metabolic gene homoeologs are involved in beta-carotene accumulation in wheat grain endosperm. We have also isolated Targeting Induced Local Lesions in Genomes (TILLING) mutants of the carotenoid metabolic gene homoeologs. Building upon the molecular knowledge and the genetic resources, we are currently determining the contribution of the key carotenoid metabolic gene homoeologs, singly and in combination, to beta-carotene accumulation in the endosperm using TILLING mutants. Our results on gene expression analysis and progress on mutant identification and characterization will be presented.

P1039: Wheat, Barley, Oat, and related

Molecular Dissection of AGPase Enzyme and its Genes in Wheat and Ten Other Species

Ritu Batra[1], Gautam Saripalli[1], Amita Mohan[2], Sachin Kumar[1], Kulvinder S. Gill[2], Harindra S. Balyan[1] and Pushpendra K. Gupta[1], (1)Ch. Charan Singh University, Meerut, India, (2)Washington State University, Pullman, WA

ADP-glucose pyrophosphorylase (AGPase) is a heterotetramer with two large subunits (LS) and two small subunits (SS). It plays a critical role in starch biosynthesis. Using the well characterized *Sh2* (LS: large subunit) and *Bt2* (SS: small subunit) genes of maize AGPase as references, true orthologs were identified in seven other monocots (*Triticum urartu*, *Aegilops tauschii*, wheat, rice, barley, sorghum and *Brachypodium*) and three dicots (*Arabidopsis*, chickpea and potato). The detailed structure, function and evolution of the genes encoding the LS and the SS among monocots and dicots were studied. The results of the present study suggested that: (i) at the DNA level, the genes controlling the SS are more conserved than those controlling the LS; the variation in both is mainly due to intron number, intron length and intron phase distribution; (ii) at protein level, the SS genes are more conserved relative to those for LS; (iii) "QTCL" motif (providing thermostability to *AGPase*) present in SS showed evolutionary differences in *AGPase* belonging to wheat 7BS, *T. urartu*, rice and sorghum, while "LGGG" motif in LS was present in all species except *T. urartu* and chickpea; (iv) expression analysis revealed downregulation of both subunits under conditions of heat and drought stress. The wheat sequences identified in the present study will be utilized to design genome specific primers. These primers will be used to amplify the three copies each of *AGPase* LS and SS genes located on homoeologous group 1 and 7 chromosomes, respectively in a set of wheat genotypes (20 heat tolerant, 20 heat sensitive and 8 moderately heat tolerant/sensitive) to identify alleles of *AGPase* LS and SS genes that may impart thermotolerance.

P1040: Wheat, Barley, Oat, and related

Fine Mapping a Major QTL Controlled Tiller Number and Plant Height Using a Wheat660K SNP Array

Tao Xu, Haiyan Wang, Xiue Wang and Jin Xiao, Nanjing Agricultural University, Nanjing, China

Tiller number and plant height are two major agronomic traits in cereal crops affecting plant architecture and grain yield. NAUH167, a mutant of common wheat landrace Wangshuibai induced by ethylmethyl sulfide (EMS) treatment, exhibits higher tiller number and reduced plant height. A stable major QTL designated *QHt.nau-2D* controlling plant height and tiller number, was mapped to the short arm of chromosome 2D flanked by markers *QHT239* and *QHT187* covering a predicted physical distance of 6.77 Mb. To further map the *QHt.nau-2D* loci, a population consisted of 6009 F_2 progeny from a cross 2011I-78 /NAUH167 was constructed. At the same time, additional molecular markers were developed to saturate the *QHt.nau*-2D region based on the Wheat660K SNP array. On the basis of Chinese Spring sequences, 53 ARMS-PCR and 18 CAPS/dCAPS markers were designed to detect the polymorphism between 2011I-78 and NAUH167. Finally, *QHt.nau-2D* was located within a genetic region of 0.5 cM between markers *QHT239* and *SNP17* spanning a 1.22 Mb physical genomic region of *Ae. tauschii* chromosome 2DS. The genetic and physical maps of *QHt.nau-2D* provide a framework for map-based cloning and this research would facilitate the characterization of plant height and tiller number in wheat.

P1041: Wheat, Barley, Oat, and related

QTL Mapping of Flag Leaf-Related Traits in Wheat (*Triticum aestivum* L.)

Jinkun Du, China Agricultural University, Beijing, China

This study aimed to advance our understanding of the genetic mechanisms underlying morphological traits of the flag leaves of wheat (*Triticum aestivum* L.). A recombinant inbred line (RIL) population derived from ND3331 and the Tibetan semi-wild wheat Zang1817 was used to identify quantitative trait loci (QTL) controlling flag leaf length (FLL), flag leaf width (FLW), flag leaf area (FLA) and flag leaf angle (FLANG). Using an available high-density simple sequence repeat (SSR) genetic linkage map, 24 putative QTL for FLL, FLW, FLA, and FLANG were detected on chromosomes 1B, 2B, 3A, 3D, 4B, 5A, 6B, 7B, and 7D. Individual QTL explained 4.2–68.52% of the phenotypic variance in different environments. Four QTL for FLL, two for FLW, four for FLA and five for FLANG were detected in at least two environments. Eighteen QTL for flag leaf-related traits originated from ND3331 alleles, and six originated from Zang1817 alleles. QTL with pleiotropic effects or multiple linked QTL were also identified on chromosomes 1B, 4B, and 5A; these are potential target regions for fine mapping and marker-assisted selection in wheat breeding programs.

Keywords: wheat; flag leaf size; flag leaf angle; QTL mapping

P1042: Wheat, Barley, Oat, and related

Improved Markers for a Pre-Harvest Sprouting QTL on Wheat Chromosome 3B.

Mark Jordan, Agriculture & Agri-Food Canada, Morden, MB, Canada and Frank M. You, Agriculture and Agri-Food Canada, Morden, MB, Canada

A QTL for sprouting index on chromosome 3B was identified in the AC Domain x RL4452 bi-parental mapping population (Cabral et al. 2014, BMC Plant Biology 14:340). The flanking markers derived from the Illumina iSelect 90K chip define a region of 185 Mbp in the wheat RefSeq 1.0 genome sequence containing over 1180 annotated high confidence genes. Additional markers are needed to narrow down this region for wheat breeding purposes as well as to eventually identify candidate genes. As a first step to identify additional markers SNPs previously validated in the KASP assay were identified by locating QTL flanking markers from the 'AC Domain' x RL4452 3B map on the 'Avalon' x 'Cadenza' map. Validated KASP markers mapped on Avalon x Cadenza in the region between QTL flanking markers were chosen (http://lgcapps.com/assays/wheat/) for genotyping and mapping on the 'AC Domain' x RL4452 population. Subsequently the full genomic region between flanking markers from the IWGSC RefSeq1.0 genome sequence was used to identify SNPs between the parental genotypes in a sequence database derived from exome capture sequencing of Canadian wheat genotypes. 1492 SNPs were identified in 288 annotated genes. The SNPs will be useful to further narrow down the QTL region and provide robust markers for selection of lines with increased tolerance to pre-harvest sprouting.

P1043: Wheat, Barley, Oat, and related

Allelic Contributions of TaPHS1 and TaMKK3 to Pre-Harvest Sprouting Resistance in Montana Spring and Winter Wheats.

Justin Vetch[1], Robert Stougaard[2], John M. Martin[3] and Mike Giroux[1], (1)Montana State University, Bozeman, MT, (2)Montana State, Kalispell, MT, (3)Montana State Univ, Bozeman, MT

Pre-harvest sprouting (PHS) is the precocious germination of grain prior to harvest which negatively impacts seed and end use quality. PHS is evaluated in two ways: visual inspection or the falling number (FN) test. A low FN value denotes starch degradation caused by PHS. This negatively impacts producer prices resulting in global losses of up to $1 billion per year. Thus, identifying wheat varieties which are PHS resistant is important. The first goal of our project was to screen Montana wheat varieties for FN and PHS tolerance. Our second goal is to investigate the contribution of allelic variation in the TaPHS1 and TaMKK3 genes which have been reported to be associated with PHS. A PHS screening method was developed and used to screen ~50 Montana grown spring and winter wheat varieties. During the PHS screening, high variability in PHS was observed in both spring and winter wheat varieties. Among MT grown winter wheats, TaPHS1 allelic variation was associated with significant PHS variation. MT grown spring wheats did not vary for the previously reported TaPHS1 resistant and susceptible alleles. Both spring and winter wheats varied for TaMKK3 alleles with those carrying the resistant allele trending lower in PHS. We hypothesize that there is unpublished allelic variation in TaPHS1 contributing to PHS variation in MT spring wheats and we are currently identifying additional TaPHS1 sequence variants. The results of this study will be used to develop PHS resistant spring and winter wheat varieties adapted to Montana.

P1044: Wheat, Barley, Oat, and related

Development of a Complete Set of Wheat-Barley Group-7 Robertsonian Translocation Chromosomes Conferring an Increased Content of ß-Glucan

Tatiana V. Danilova[1], Bernd Friebe[1], Bikram Gill[1], Jesse Poland[2] and Eric W. Jackson[3], (1)Kansas State University, Manhattan, KS, (2)Department of Plant Pathology, Kansas State University, Manhattan, KS, (3)General Mills, Minneapolis, MN

Many valuable genes for agronomic performance, disease resistance and increased yield have been transferred from relative species to wheat (*Triticum aestivum* L.) through whole-arm Robertsonian translocations (RobT). Although of a great value, the sets of available translocations from barley (*Hordeum vulgare* L.) are limited. Here we present the production of a complete set of six compensating RobT chromosomes involving barley chromosome 7H and three group-7 chromosomes of wheat. The barley group-7 long arm RobTs had a higher grain ß-glucan content compared to the wheat control. The ß-glucan levels varied depending on the temperature and were higher under hot conditions. Implicated in this increase, the barley cellulose synthase-like F6 gene (*CslF6*) responsible for ß-glucan synthesis was physically mapped near the centromere in the long arm of barley chromosome 7H. Likewise, wheat *CslF6* homoeologs were mapped near the centromere in the long arms of all group-7 wheat chromosomes. With the set of novel wheat-barley translocations, we demonstrate a valuable increase of ß-glucan, along with a resource of genetic stocks that are likely to carry many other important genes from barley into wheat.

P1045: Wheat, Barley, Oat, and related

Genome-Wide Association Mapping of Phenotype Traits Related to Phosphorus Use Efficiency in Synthetic Hexaploid Wheat

Emily K. Gordon[1], Sateesh Kagale[2], Mina Kaviani[1], Istvan Rajcan[1], Dave Hooker[3] and Alireza Navabi[1], (1)University of Guelph, Guelph, ON, Canada, (2)National Research Council Canada, Saskatoon, SK, Canada, (3)University of Guelph, Ridgetown, ON, Canada

Phosphorus (P) is a non-renewable macronutrient required for many plant processes, where P fertilizer uptake rates are typically <30% of what is applied. Not only are current application rates unsustainable, fertilizer application from agriculture production is a major source of P loading into water bodies around the world. Breeding for more P efficient crops may improve the P balance efficiency of the entire cropping system by preventing dissolved P from being lost, exported and accumulated in the field. As a major cereal crop worldwide, wheat (*Triticum aestivum* L.) yields must continue to increase in order to meet expected food demands under lower P inputs. Wheat has the potential to be improved in terms of its genetic diversity for P use efficiency. To investigate this, a panel of 194 synthetic hexaploid wheat (SHW) derived accessions from the International Maize and Wheat Improvement Centre (CIMMYT) were genotyped with the Illumina iSelect 90k single nucleotide polymorphism (SNP) chip and phenotyped under 0 and 60 kg P ha^{-1} in three trials over two years in Ontario, Canada. We identified genomic regions on chromosomes 2B and 5B associated with yield under high P. Additionally, a region on chromosome 3B showed association with P deficiency tolerance. Further identification of genomic regions associated with the phenotypic traits analysized in this study will be presented.

P1046: Wheat, Barley, Oat, and related

Genome-Wide Association Mapping Reveals Major Genomic Regions for Grain Zinc Concentration in Wheat

Govindan Velu, CIMMYT, Texcoco, Mexico

Bread wheat is a major staple providing 20% of dietary energy and major source of protein and essential micronutrients such as iron (Fe) and zinc (Zn) for world's population. About two billion people are deficient in some essential micronutrients, including the Zn and Fe deficiency. The magnitude of Fe and Zn deficiency is particularly severe among children and women. To close nutrition gaps in rural households of remote areas, development and dissemination of high-yielding and nutrient-rich wheat varieties offers a cost-effective and sustainable solution. Breeding for enhanced Zn concentration in wheat was initiated by crossing high Zn progenitors such as synthetic hexaploid wheats, T. dicoccum, T. spelta and landraces. These crosses have resulted in several wheat varieties with competitive yields and enhanced grain Zn which were adapted by thousands of small-holder farmers in South Asia and Africa. Here we report one of our genome-wide association studies (GWAS) using the wheat 90K genotyping assay, characterized in 330 bread wheat cultivars. The diverse HarvestPlus Association Mapping (HPAM) panel was phenotyped in a range of environments in India and in Mexico. The GWAS analysis revealed more than 30 marker-trait associations (MTA) for grain Zn in wheat. Interestingly two large effect QTL regions were found on 2 and 7 chromosome groups. Moreover, 3 to 4 known candidate genes associated with Zn homeostasis and metal transporter genes were mapped near these QTL regions. The markers and associated candidate genes identified in this study are being validated in new biparental mapping populations and breeding materials.

P1047: Wheat, Barley, Oat, and related

Wheat miR9678 Controls Seed Germination By Generating Phased ta-siRNAs and Modulating Abscisic Acid/Gibberellin Signaling

Yingyin Yao, China Agricultural University, Beijing, China
Seed germination is important for wheat yield and quality. However, our knowledge of mechanisms regulating seed germination in wheat remains limited. In this study, we found microRNA9678 (miR9678) is specifically expressed in the scutellum of developing and germinating wheat seeds. Overexpression of miR9678 delays germination and improves resistance to pre-harvest sprouting (PHS) in wheat; miR9678 silencing enhances germination rates. miR9678 triggers phased *trans*-acting small interfering RNAs (ta-siRNAs) by cleaving the long non-coding RNA, and ta-siRNAs also delay seed germination. In addition, miR9678 overexpression also reduces bioactive gibberellin (GA) levels through a ta-siRNAs-independent mechanism. Finally, abscisic acid (ABA) signaling proteins bind the promoter of miR9678 precursor and activate its expression, indicating miR9678 regulates germination by modulating the GA/ABA signaling.
Keywords: wheat, miRNA, seed germination, phased ta-siRNAs, gibberellin, long non-coding RNA

P1048: Wheat, Barley, Oat, and related

Genome-Wide Analysis of MIKC-Type MADS-Box Genes in Wheat: A Primer for Crop Improvement

Alice Kennedy, Susanne Schilling, Sirui Pan and **Rainer Melzer**, University College Dublin, Dublin, Ireland
MIKC-type MADS-box genes encode transcription factors with prominent roles in plant development. They constitute key regulators of flowering time, floral organ identity, seed and fruit development. MADS-box genes have also been the target of domestication processes in numerous plant species. Understanding the function and evolution of MADS-box genes in crops is therefore of considerable interest for future crop improvement programs.
Here, we present a genome-wide analysis of the MIKC-type MADS-box gene family from wheat. We identified more than 200 MIKC-type MADS-box genes, considerably more than in other monocots, partly due to the hexaploid nature of wheat. MIKC* genes as well as representatives from 15 distinct MIKC^c subfamilies were identified. Our preliminary analyses indicate that some MADS-box gene subfamilies (e.g. AGL17-like genes) expanded considerably in wheat whereas others (e.g. AGL6-like genes) have relatively few members as compared to other moncots. Using *in silico* analyses we deduced a relatively strong conservation of expression pattern within each subfamily, indicating functional similarity among closely related homologs.
We will use the identified MADS-box genes to screen for sequence variation among different wheat lines. This will provide a starting point to reveal how allelic variation in MADS-box genes may affect agronomically important traits in wheat.

P1049: Wheat, Barley, Oat, and related

Wheat WRKY Gene *TaWRKY51* Plays Positive Roles in Drought Stress

Xingbei Liu, Mei Zheng, Zhongfu Ni, Qixin Sun and **Zhaorong Hu**, China Agricultural University, Beijing, China
WRKY-type transcription factors are involved in multiple aspects of plant growth, development and stress response. WRKY genes have been found to be responsive to abiotic stresses; however, their roles in abiotic stress tolerance are largely unknown especially in crops. Wheat (*Triticum aestivum* L.) is one of the major crops largely cultivated and consumed all over the world. The molecular mechanism of the abiotic stress response in wheat is largely unclear. We previously identified multiple stress responsive WRKY genes from wheat. Here, we further characterized the roles of one of these genes, *TaWRKY51*, in abiotic stress tolerance. *TaWRKY51* expression was increased by various abiotic stresses. Over-expression and RNAi analysis demonstrated that *TaWRKY51* improves drought tolerance in transgenic wheat lines. Measurement of physiological parameters, including chlorophyll and proline contents, supported this conclusion. TaWRKY51 enhanced expressions of *NCEDs* and *DREBs* genes. TaWRKY51 protein may regulate the downstream genes through direct binding to the gene promoter or via indirect mechanism. Manipulation of *TaWRKY51* in wheat or other crops should improve their performance under drought stress conditions.

P1050: Wheat, Barley, Oat, and related

Genome-Wide Association Study of Winter Bread Wheat (*Triticum aestivum* L.) in Response to Drought in a Multi-Environment European Network.

Gaëtan Touzy[1], Fred E. Van Eeuwijk[2], Renaud Rincent[3], Willem Kruijer[4], Stéphane Lafarge[5], Katia Beauchene[1], Jacques Le Gouis[3], Sébastien Praud[6] and The BreedWheat Consortium, (1)ARVALIS Institut du Végétal, VILLIERS LE BACLE, France, (2)Wageningen University & Research, WAGENINGEN, Netherlands, (3)INRA, Clermont-Ferrand, France, (4)Wageningen University & Research - Biometris, Wageningen, Netherlands, (5)Biogemma, Chappes, France, (6)LIMAGRAIN, CHAPPES, France
Keywords: Drought stress, Genome-wide association mapping, GWAS, Genotype-by-environment interaction, multi-trait multi-environment association mapping.
Drought is one of the main abiotic stresses limiting wheat (*Triticum aestivum L.*) growth and productivity around the world. Many climate-based simulations have predicted an increase in the frequency and intensity of this abiotic stress. The delivery of new high yielding and stress-tolerant cultivars is now necessary and requires an improved understanding of the basis of the physiological and genetic response to drought. A panel of 220 European elite cultivars was evaluated in 32 field experiments. Grain yield and yield components were scored for each trial. A crop model was run with detailed climatic data and soil water status, to identify the timing, intensity and history of stress for each combination of genotype/trial. Cultivars were genotyped with the TaBW420K chip. This dataset gives us the opportunity for a detailed study of genetic by environmental interactions.
Three scenarios of water deficit have been identified in this trial network. The grain yield loss in the two stressed scenarios was between 7 to 12% when compared to the non-stressed scenarios. A large genetic variability of grain yield was identified, with a genotypic variation affecting the mean by ± 15%. In the same way, GxE interactions affected the grain yield mean by 12.5%. GWAS were performed using multi-environment mixed models. Several QTLs were identified in the different stress scenarios, the allelic effects of these QTLs have been related to the environmental co-variables. Methods and results will be discussed especially those regarding the impacts of QTLxE interactions on grain yield and components and grain yield.

Acknowledgements: we thank the CIFRE, BreedWheat (ANR-10-BTBR-0003) and for financial support and French breeding companies for conducting trials.

P1051: Wheat, Barley, Oat, and related

Evaluation of Winter-Survival of Winter Wheat By Drone-Generated Multi-Spectral Imagery Identified Two Quantitative Trait Loci on Chromosome 5A

Yi Chen, Mina Kaviani, Harwinder Singh Sidhu and Alireza Navabi, University of Guelph, Guelph, ON, Canada

The harsh and unpredictable winters in the high latitude regions of the northern hemisphere often leads to high risk of winterkill for winter wheat (*Triticum aestivum* L.). One of the key traits that influence winter-survival is the timing of the transition from vegetative to reproductive stage, as wheat loses cold tolerance after the transition. The goal of this research is to investigate the genetic basis of winter-survival in Canadian winter wheat and to identify the combination of key candidate gene alleles that is optimal for high-latitude winter wheat. The Canadian Winter Wheat Diversity Panel (CWWDP), which includes 450 winter wheat genotypes with various levels of winter-hardiness, was phenotyped under field conditions in Ontario, Canada. Normalized difference vegetation index (NDVI) was extracted from multi-spectral imagery captured by unmanned aerial vehicle (UAV) as a measure of winter-survival. The diversity panel was genotyped with the 90K Illumina SNP chip and additionally for allelic variation at the candidate gene loci that have demonstrated significant effect on flowering time and cold tolerance. This included the major vernalization gene loci on group 5 chromosomes (*VRN-A1, VRN-B1,* and *Vrn-D1*), C-Repeat Binding Factor (*CBF*)-*12* and -*15* on chromosome 5A and photoperiod response loci on chromosome 5D (*PPD-D1*) and 5A (*PPD-A1*). Genome-wide association studies identified two major quantitative trait loci on chromosome 5A, which correspond to the frost resistance loci *Fr-A1* and *Fr-A2*. This result is consistent with previous reports on the role of chromosome 5A in winter-survival and showcased the potential of UAV-based phenotyping in genetics research.

P1052: Wheat, Barley, Oat, and related

Technologies to Increase Genetic Potential Productivity Under Abiotic Stress for Winter Wheat and Barley Crops, in Climatic Condition from Braila Plain, Romania

Daniela Trifan, Agricultural Research and Development Station, Braila, Romania; BRAICOOP Agricultural Cooperative - Research Department, Braila, Romania

Each agricultural crop has a theoretical genetic potential, which is represented by the production quantity and quality obtained by the variety of culture, in perfect condition. But perfect conditions for culture there are in few places, especially lately, when seen increasingly accelerated phenomenon of climate change. Increasing theoretical genetic potential for productivity was achieved for each species grown in many years of genetic research, both in the laboratory and in the field experiences. Every year is putting out more and more performing varieties with increased resistance to drought, heat, pests and diseases etc. But genetic potential to the maximum occurs when growing conditions are as close genetically programmed requirements. The abiotic stress can significantly reduces the genetic productivity potential from the very early stages of germination and the vegetation, if the conditions are not fulfilled optimum microclimate (humidity, temperature, nutrients, absence of pests and diseases, etc.). On the other hand, even if the plant grew and developed normally, but is attacked by diseases or pests in a phase of vegetation close to reproduction, this can significantly reduce production unless corrective measures and effective protection of agricultural crops. This paper presents the results about monitoring of genetic potential at some winter wheat and barley varieties, and the results of agrophytotechnical methods to increase genetic potential on production and product quality. The experiments during in the period 2012 - 2017, at Agricultural Research and Development Station of Braila, Romania, by comparing the genetic production potential of some varieties of wheat and barley under different densities and dates of sowing, different fertilization (chemical and biological), and the application of plant biostimulators, also.

P1053: Wheat, Barley, Oat, and related

RNA-Seq Analysis Reveals Jasmonates Related Pathways Associate with Salinity Tolerance in Wheat

Qiaoling Luo, Bin Li, Hongwei Li, Fangpu Han, Zhensheng Li and **Qi Zheng**, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

To explore the salt tolerance mechanism of wheat, we carried out RNA sequencing with 12 samples from three seedling tissues of salt-tolerant variety Xiaoyan 60 and high-yielding variety Zhongmai 175 under the salt treatment and the control. After analysis of different expression, 703, 979, 1197 differentially expressed genes (DEGs) were found respectively in new leaves, old leaves and root in Xiaoyan 60 when compared the salt treatment and the control, while the corresponding DEGs number in Zhongmai 175 were 613, 1401 and 1301. Further analysis demonstrated many DEGs were related with salt tolerance. Gene Ontology (GO) analysis showed the term “fatty acid biosynthesis process” was significantly enriched in new and old leaves of Xiaoyan 60, concurrently, the KEGG pathways “linoleic acid metabolism” and “alpha-linolenic acid metabolism” were also enriched. And most DEGs in these processes were up regulated, which indicated the level of jasmonate could be improved because the synthesis of jasmonate (JA) was through “alpha-linolenic acid metabolism”. In root tissue of Xiaoyan 60, the most significantly enriched KEGG pathway was “glucosinolate biosynthesis”, which could be induced by JAs. Differently, the most significantly enriched GO terms in the new and old leaves of Zhongmai 175 were “response to red or far red light” and “cellular response to starvation”. And similarly, the KEGG pathway “photosynthesis – antenna proteins” was also significantly enriched. Further analysis demonstrated that almost all the DEGs in these terms or pathways in Zhongmai 175 were down regulated, which manifested that the photosynthesis system may be damaged in Zhongmai 175, especially in the old leaves.These results indicate the jasmonates (JAs) related signal pathways may play a vital role in the salt tolerance of Xiaoyan 60. Inversely, the effects of JAs related pathways may be weaker in Zhongmai 175, and the photosynthesis system is destroyed due to the salinity stress.

P1054: Wheat, Barley, Oat, and related

Genome-Wide Association Mapping for Seedling Heat Tolerance in Winter Wheat

Frank Maulana[1], Joshua D. Anderson[1], Habtamu Tamir[1], Tadele Kumssa[2] and Xuefeng Ma[3], (1)Noble Research Institute, LLC, Ardmore, OK, (2)Noble Research Institute,LLC, Ardmore, OK, (3)Noble Research Institute, LLC., Ardmore, OK
Heat stress at seedling stage is one of the most common issues of winter wheat in a dual-purpose management system in the southern Plains. However, the genetic mechanism underlying seedling heat tolerance in wheat is not well studied. To dissect the genetic basis of this trait, we conducted a genome-wide association mapping study (GWAS) using 200 hard red winter wheat lines from the Triticeae Coordinated Agricultural Project (TCAP), genotyped using the wheat *i*select 90K SNP genotyping array. The plants were initially planted under optimal temperature in growth chambers. At three leaf stage, plants were subjected to two temperature regimes, high temperature (40/35ºC day/night) as heat stress treatment, and optimal growth temperature (25/20ºC day/night) as control for 14 days. Data were collected on leaf chlorophyll content (LCC), shoot length (SL), number of leaves (LN), and percent of seedling recovery (PSR) under optimal growth temperature following the heat stress treatment. GWAS was performed using mixed linear model (MLM). Significant marker-trait associations were found on all the traits under both optimal and heat-stressed growth conditions. In addition, heat stress responding marker-trait associations were also detected. Once validated, these SNPs will be used in marker-assisted selection of seedling heat tolerance in wheat.

P1055: Wheat, Barley, Oat, and related
Discovering Superior Alleles Underlying Drought and Heat Tolerance in Common Wheat
Ruilian Jing, Xinguo Mao, Jingyi Wang and Xiaoping Chang, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China
Environmental stresses, drought and heat, especial the stresses in grain filling period are the primary causes of grain yield losses in wheat (*Triticum aestivum* L.). Abundant wheat germplasm resources with tolerance to drought (DT) and heat (HT) have been identified. They are important gene resources for wheat improvement. However, very little is known about these resources, such as what tolerant-genes these germplasm possess, in which accessions have the superior alleles underlying the tolerances, and how to use these alleles effectively. Our researches showed that some genes involved in the DT and HT, also contributed to the yield-related traits. Here, we dissect the functions of gene *TaSnRK2* (sucrose non-fermenting1-related protein kinase 2), *TaPP2A* (protein phosphatase 2A), and *TaSPL* (squamosa-promoter binding protein, SBP-box). *TaSnRK2* and *TaPP2A* not only responded to abiotic stress, but also significantly associated with 1000-grain weight under terminal drought and heat stress. Two of the *SPL* members, *TaSPL20* and *TaSPL21* associated with 1000-grain weight and plant height in multi-environment. According to the functional analysis of naturally occurring variants, favorable alleles/haplotypes were identified with the perfect markers. They increased 1000-grain weight, and reduced plant height in well-watered, drought, and heat environments. Our study suggests that during domestication and breeding of wheat in China, superior alleles of each gene were selected and exploited to varying degrees due to their large effects on the yield-related traits.

P1056: Wheat, Barley, Oat, and related
"Am Hidden, Feeling Hot! Do Something!" – Heat Stress on Wheat Roots
Sundaravelpandian Kalaipandian, CSIRO (Agriculture and Food), Brisbane, Queensland, Australia, Anne L. Rae, CSIRO (Agriculture and Food), St Lucia, Australia, Gangping Xue, CSIRO (Agriculture and Food), Australia and C. Lynne McIntyre, CSIRO (Agriculture and Food), Brisbane, QLD, Australia
High temperature is a major threat to plant productivity due to climate change. Often the hidden-half of the plant is more sensitive to heat stress than the above ground parts. Heat stress affects the roots by limiting water and nutrient uptake, which in turn affect shoot water demand and photosynthesis. However, the molecular mechanism of root responses to heat stress is poorly understood. Recently, we have studied the role of TaHsfC2a gene in wheat. Overexpression of TaHsfC2a-B in transgenic wheat plants increased survival rate to about 90% while only 15% of wild-type plants survived after heat treatment at 43°C. Interestingly, we observed that the shoots were drying but the roots were intact after heat treatment in the transgenic plants, which contributed to recovery of the shoots, however all parts of the wild-type plants died after heat treatment. Reactive oxygen species (ROS) was shown to have a major role in abiotic stresses including heat stress. We found that the transgenic plant roots accumulate very low hydrogen peroxide (H_2O_2) when compared with wild-type plant roots. To understand the molecular mechanisms underlying the heat stress and ROS in the roots, the transcriptome of TaHsfC2a transgenic and wild-type roots are being studied by using RNA-sequencing. In addition, we found that TaHsfC2a was markedly up-regulated under drought and abscisic acid treatment, and we have also identified the potential targets of this gene (TaHSP70d and TaGalSyn) and confirmed through transactivation studies. Our study will identify candidate genes to develop heat resistant varieties in wheat.

P1057: Wheat, Barley, Oat, and related
Leveraging the Root Angle QTLome to Enhance Climate Resilience in Wheat
Marco Maccaferri, DipSA, Department of Agricultural Science, University of Bologna, Bologna, Italy, Danara Ormanbekova, Department of Agricultural Sciences, Bologna, Italy; DipSA - University of Bologna, Bologna, Italy, Ghasemali Nazemi, Department of Agriculture, , Haji abad, Iran (Islamic Republic of) and **Roberto Tuberosa**, Department of Agricultural Sciences, University of Bologna, Bologna, Italy
Optimisation of root system architecture (RSA) is an important objective for the sustainability of durum wheat grown under drought-stressed conditions. In the present study, linkage and association mapping (AM) for RSA evaluated at the seedling stage evidenced 20 clusters of quantitative trait loci (QTLs) for root length and number as well as 30 QTLs for root growth angle (RGA). The most divergent RGA phenotypes observed by seminal root screening were validated by root phenotyping of field-grown adult plants. QTL analysis of RSA and grain yield data indicates RGA as a valuable target to enhance grain yield and yield stability across different soil moisture regimes (Maccaferri et al. 2016). Based on their relative additive effects, allelic distribution in the AM panel and co-location with QTLs for yield, eight RGA QTLs have been prioritised in terms of breeding interest and value. These QTLs were investigated for gene content based on the chromosomal pseudomolecules of Chinese Spring *T. aestivum* and the TriAnnot v4.3 gene prediction and annotation pipeline and the Zavitan *T. dicoccoides* genome assembly (Avni et al. 2017). The chromosome regions contained 25 to 242 predicted genes (123 on average). In six RGA QTLs, from one to four gene annotations were involved in auxin pathways. The comparison between the *T. aestivum* and *T. dicoccoides* gene content

indicates the high quality of the *T. dicoccoides* assembly and its usefulness to identify candidates to explore the polymorphism and the structural variation of drought-related genes present in the A and B wheat genomes.

P1058: Wheat, Barley, Oat, and related

Functional Study of TabZIP15 in Regulation of Wheat Abiotic Stress Tolerances

Lichao Zhang[1], Xiuying Kong[2] and Yiqian Zhai[2], (1)Institute of Crop Sciences, Beijing, China, (2)Institute of Crop Sciences, China

bZIP transcription factors are one of the most important transcription factor families which play important roles in response to biotic and abiotic stresses. However, few studies of the functions of bZIP transcription factors in regulation of abiotic stresses tolerance have been done in wheat.

TabZIP15 encoded a bZIP transcription factor of C subfamily, which was mapped on the wheat chromosome 7DL. *TabZIP15* was induced by salt, PEG, cold stresses and exogenous ABA treatment. The protein encoded by *TabZIP15* was localized in the nucleus through transient expressed in tobacco epidermal cells, and possessed transcription activation activity in yeast with an N-terminal transcriptional activation domain. Overexpression of *TabZIP15* improved the drought and freeze tolerance of transgenic *Arabidopsis* plants. Yeast one-hybrid experiments showed that TabZIP15 transcription factor can bind to ABRE cis-acting elements. Yeast library screening experiments and luciferase complementation assay (LCI) showed that TabZIP15 can interact with enolase TaENO-b, indicating that TabZIP15 may regulate abiotic stress tolerance through glycolysis and gluconeogenesis pathway.

P1059: Wheat, Barley, Oat, and related

A Micro-and Macrophenomics Framework for Cereal-Pathogen Interactions

Patrick Schweizer, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany

A Micro- and Macrophenomics Framework for Cereal-Pathogen Interactions

The precise assessment of quantitative host resistance against plant pathogens is challenging because the trait is usually controlled by a number of genes and loci with small to moderate effects. Therefore, phenotyping by accurate and high throughput methods is required including microscopic approaches to characterize early stages of plant-pathogen interactions. To meet this challenge we have developed a "microphenomics" platform that includes automated microscopy and image analysis, and that can be combined with high-throughput DNA cloning and single-cell transformation protocols for gene functional studies. The parameters measured range from initial haustorium establishment over early colony formation and -size to coverage of plant surface by mature sporulating colonies. These phenomic tools cover the complete asexual reproduction cycle of powdery mildew fungi thereby allowing strictly quantitative assessment of disease development and host responses. All these tools are part of the Micro- and Macrophenomics framework and can be used in answering questions about the role of specific genes and genotypes in plant disease resistance or susceptibility.

Title: A Micro- and Macrophenomics Framework for Cereal-Pathogen Interactions

Submitter's E-mail Address: schweiz@ipk-gatersleben.de

Session Selection: Wheat, barley, rye, oats and related

P1060: Wheat, Barley, Oat, and related

A Transcriptome Analysis of Genes Involved in the Production of Reactive Oxygen Species By P. triticina during Its Infection on Wheat.

Xiben Wang[1], M. Z. Che[2], Hala Badr Khalil[3], Brent McCallum[1], Guus Bakkeren[4] and Barry J. Saville[5], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)Morden Research and Development Centre, Agriculture and Agri-Food Canda, Morden, MB, Canada, (3)Agriculture and Agri-food Canada, Summerland, BC, Canada, (4)Agriculture and Agri-Food Canada, Summerland, BC, Canada, (5)Trent University, Peterborough, ON, Canada

Reactive oxygen species (ROS) play an important role during host and pathogen interactions and are often an indication of induced host defense responses. The importance of these radicals for pathogenesis of the obligate biotrophic fungus, *Puccinia triticina* (*Pt*), has not been investigated. In this study, we demonstrate that *Pt* generates ROS, including superoxide, H_20_2 and hydroxyl radicals, during its infection of wheat. Through pharmacological inhibition, we show that ROS are critical for both *Pt* urediniospore germination and intercellular growth. A *Pt* genome-wide screening identified 291 putative genes associated with general redox homeostasis and the search in RNA-seq data sets representing *Pt* urediniospore germination, early and late infection stages on susceptible wheat cultivar Thatcher, found 37 genes annotated to encode known products related to oxidative stress responses. We identified two canonical *Pt* genes encoding NADPH oxidases (*PtnoxA* and *PtnoxB*), as well as a regulatory gene, *PtnoxR*. Real time qPCR analysis showed that all three genes were differentially regulated during urediniospore germination and infection on wheat.

P1061: Wheat, Barley, Oat, and related

Wheat Mutants with Reduced *Puccinia triticina* Infection

John Fellers, USDA ARS, Manhattan, KS

Wheat and *Puccinia triticina*, the fungus causing leaf rust, have a biotrophic relationship that has evolved over time. After a urediniospore lands on the leaf, it will germinate, form an appresoria over the stomate, insert a hyphal tube, find a mesophyll cell, and produces a haustoria which invaginates the host cell plasma membrane. The haustoria will begin to secrete effectors garnering host nutrients, squelching host defenses, and will entice host enzymes and pathways to work for the fungus. Recent cloning of broad-spectrum resistance genes suggests that stable resistance may be found by upsetting the biotrophic interaction between rust and wheat. To identify wheat genes needed by the pathogen, EMS was used to mutate the spring cultivar Thatcher. M_2 lines were screened with *P. triticina* race BBBD, and lines were identified with a reduction in rust infectivity. Lines were taken to the M_6 and evaluated in the field under natural, mixed race infections, and 25 lines were identified. Phenotypes included little or no pustules, race specific-like resistance, or reduction in pustule size. These lines were backcrossed to Thatcher and resistant and susceptible F_2 lines were pooled and their RNA and DNA sequenced. Phenotypes segregated in a manor consistent

with a single recessive gene. Comparisons back to the wheat genome will be made and genes will be identified and reported that are associated with the traits.

P1062: Wheat, Barley, Oat, and related

Wheat Leaf Rust Resistance Gene from Marquis Wheat

Brent McCallum[1], Mingzhe Che[2], Colin Hiebert[1] and Curt McCartney[1], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)Department of Plant Pathology, China Agricultural University, Beijing, China

Marquis wheat, released in 1911, was one of the most widely grown cultivars in Canada and the north-central USA. It was susceptible to all leaf rust (*Puccinia triticina*) isolates tested, up to the emergence of a group of races in the early 2000s, predominantly TDBG. Marquis had an unusual mesothetic resistance phenotype when inoculated with TDBG. To characterize this resistance the Marquis backcross line RL6071was crossed with a leaf rust resistant accession from the Kyoto University wheat germplasm collection KU168-2 to create a doubled haploid population. Seedling resistance from RL6071 was inherited as a single resistance gene that mapped to chromosome 7BL. Tightly linked molecular markers, along with seedling leaf rust testing and pedigree analysis revealed that this gene, temporarily named *LrMar*, was present in Marquis, Red Fife and a number of cultivars derived from Red Fife, such as White Fife, Percy and Renfrew. The same group of races that were avirulent to *LrMar* were also avirulent to *LrCen*, previously mapped to 7AL, with a similar mesothetic infection type. Both genes are only effective against this small group of *P. triticina* isolates, are ineffective in conditioning field resistance against the broader Canadian population, and neither were detected prior to the emergence of these races. These could be homeologous resistance genes based on their respective positions on chromosomes 7BL and 7AL, and phenotypic similarities.

P1063: Wheat, Barley, Oat, and related

Accelerated Cloning and Characterization of the Wheat Adult Plant Resistance Gene *Lr68*

Sreya Ghosh[1], Burkhard Steuernagel[1], Caixia Lan[2], Miroslava Karafiátová[3], Ksenia Krasileva[4], Jaroslav Doležel[3], Evans Lagudah[5], Ravi P. Singh[2] and Brande Wulff[1], (1)John Innes Centre, Norwich, United Kingdom, (2)CIMMYT, Mexico City, Mexico, (3)Institute of Experimental Botany, Olomouc, Czech Republic, (4)Earlham Institute, Norwich, United Kingdom, (5)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

Adult Plant Resistance (APR) genes are broad-spectrum, partial resistance genes that can contribute to sustainable control of wheat rust diseases. However, a lack of precise molecular markers complicates their characterization and practical use in breeding programmes. At the same time, the long generation time of wheat has become a limiting factor for breeders to respond quickly to an outbreak.

As the APRs cloned so far do not belong to any common gene family, it is not possible to use general features of these identified APRs to conduct biased searches for novel APRs. This project aims to rapidly clone the recently discovered APR gene Lr68 (Leaf Rust 68) using an unbiased gene isolation technique called MutChromSeq, which combines chromosome flow-sorting and mutational genomics, and is independent of fine mapping. It also aims to combine marker-assisted selection with accelerated generation advancement ("speed breeding") for rapid germplasm structuring and field performance evaluation. Cloning APRs allows breeders to trace genes cheaply and quickly using gene-specific markers, enabling them to build effective and durable resistance gene pyramids. It also allows us to elucidate any common mechanism of action they have, helping researchers and breeders understand better the basis of their durable resistance.

P1064: Wheat, Barley, Oat, and related

Genetic Mapping of Leaf Rust Resistance in the Tetraploid Wheat Cross Strongfield/Blackbird

Xiangyu Pei[1,2], Brent McCallum[1], Colin Hiebert[1], Anita Brule-Babel[2], Ron E. Knox[3], Yuefeng Ruan[4], Curtis J Pozniak[5] and Curt McCartney[1], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)University of Manitoba, Winnipeg, MB, Canada, (3)Agriculture and Agri-Food Canada, SCRDC, Swift Current, SK, Canada, (4)AAFC-SPARC, Swift Current, SK, Canada, (5)University of Saskatchewan, Saskatoon, SK, Canada

Leaf rust, caused by *Puccinia triticina* Eriks. (*Pt*), is an economically important disease of wheat worldwide. Deploying wheat cultivars with effective leaf rust resistance (Lr) genes is an efficient method for disease management. The genetic basis of leaf rust resistance was studied a doubled haploid (DH) population of the cross Strongfield/Blackbird. Strongfield is a widely grown durum wheat variety (*Triticum turgidum* var. *durum* L.; genome AABB) in Canada, which was developed at Agriculture and Agri-Food Canada, Swift Current. Strongfield is highly resistant to *Pt* in Canada. Blackbird (*Triticum carthlicum*; genome AABB) is susceptible to *Pt* at the seedling stage but possesses partial adult plant resistance. The genetic basis of leaf rust resistance was studied in a doubled haploid (DH) population of the cross Strongfield/Blackbird which was previously genotyped with SSR markers and the 90K wheat Infinium SNP array. Four QTLs were found on chromosomes 1B, 2B, 3A, and 3B based analysis of leaf rust reaction from inoculated field nurseries in 2016 and 2017. This population was then screened for leaf rust resistance with multiple races at the seedling stage indoors . One Lr gene was identified on chromosome 3A, mapping to the same location as the 3A QTL detected with the field leaf rust data.

Submitter's Email Address: brucepei1025@gmail.com

P1065: Wheat, Barley, Oat, and related

Evidence for the *Lr46* Leaf Rust Resistance Gene in the Wheat Cultivar Carberry

Ron E. Knox[1], Firdissa Bokore[1], Richard Cuthbert[1], Brent McCallum[2], Ron DePauw[3], Colin Hiebert[2], Heather L. Campbell[1], Brad Meyer[1], Tom Fetch[4], Yuefeng Ruan[1], Andrew G. Sharpe[5] and Vijai Bhadauria[1], (1)Agriculture and Agri-Food Canada, SCRDC, Swift Current, SK, Canada, (2)Agriculture and Agri-Food Canada, Morden, MB, Canada, (3)Advancing Wheat Technology, Swift Current, SK, Canada, (4)Agriculture and Agri-Food Canada, Brandon, MB, Canada, (5)Global Institute for Food Security (U of S), Saskatoon, SK, Canada

Stacking and deployment of pleiotropic genes for resistance to multiple fungal diseases in wheat variety development is expected to increase the durability of resistance. To achieve this gene stacking objective through molecular breeding, an understanding of which genes currently exist in adapted germplasm is necessary. The cultivar Carberry is a popular hard red spring wheat variety in Canada with good rust resistance.

Pedigree, phenotype, and DNA marker evidence suggested Carberry possesses the leaf rust resistance gene *Lr46*. *Lr46* is a slow rusting adult plant resistance gene located on chromosome 1B that provides resistance against leaf rust and other diseases such as stripe rust, stem rust, and powdery mildew. We undertook an investigation to test the hypothesis that Carberry possesses *Lr46*. A doubled haploid population comprising 297 lines was developed from the F_1 of a cross of Carberry with the universally leaf rust susceptible cultivar Thatcher. The population was evaluated for leaf rust reaction in four field nursery environments: near Swift Current SK from 2014 to 2016, and Morden MB in 2016. The population was also assessed for stem rust response in 2014 and 2016 near Swift Current. The population was genotyped using the 90K Infinium iSelect assay and following linkage map construction with JoinMap 4.1, 5457 markers were used for quantitative trait locus (QTL) analysis using MapQTL 6. Two QTL for leaf rust resistance were identified from Carberry on chromosome 1B, one of which was coincident with a stem rust resistance QTL that mapped to the location of *Lr46*.

P1066: Wheat, Barley, Oat, and related

Development of Diagnostic Markers for the Detection of Functional and Non-Functional Alleles of *Yr15*

Tzion Fahima[1], Valentina Klymiuk[1], Andrii Fatiukha[1], Lin Huang[1], Elitsur Yaniv[2], Dina Raats[1], Alan H. Schulman[2], Jorge Dubcovsky[3] and Abraham B. Korol[1], (1)Institute of Evolution and the Department of Evolutionary and Environmental Biology, Faculty of Science and Science Education, University of Haifa, Haifa, Israel, (2)LUKE & University of Helsinki, Helsinki, Finland, (3)University of California, Davis, CA

Stripe rust, caused by the fungus *Puccinia striiformis* f.s. *tritici* (*Pst*), is a destructive disease of wheat globally. Depletion of effective resistance to *Pst* in cultivated wheat has led to search for new resistance genes in the wild relatives of wheat. One of the most promising genes conferring broad-spectrum resistance to stripe rust is *Yr15*, derived from wild emmer wheat (*Triticum dicoccoides*) accession G25. *Yr15*, mapped on chromosome arm 1BS, has recently been cloned by our consortium and designated as *Wheat Tandem Kinase 1* (*WTK1*). We found *wtk1* susceptible alleles in most 274 tested durum, bread, and wild emmer wheat lines. Out of 69 tested durum and bread wheat cultivars and lines, only 33 *Yr15* introgression lines contained the functional allele (*Wtk1*) from G25 and were resistant to *Pst*. The remaining 36 susceptible lines carried non-functional alleles (*wtk1*), which included insertions of large transposable elements that resulted in changes in reading frame. Development of reliable molecular markers can facilitate the introgression of *Yr15* into new varieties via marker-assisted selection. Diagnostic markers designed based on the polymorphism between the *WTK1* alleles are preferred in order to avoid negative linkage drag. Therefore, we have designed highthroughput co-dominant KASP markers that can differentiate between the functional (*Wtk1*) and all known non-functional (*wtk1)* alleles, and can be used in breeding programs for development of modern cultivars with high resistance to stripe rust.

P1067: Wheat, Barley, Oat, and related

Association Mapping of Stem Rust in Minnesota Spring Wheat Lines

Cyrus Kimani Ndung'u, University of Minnesota, St Paul, MN

Stem rust caused by *Puccinia graminis* f. sp. *tritici,* especially the Ug99 (TTKSK) race, is a serious threat to wheat production around the world and can cause up to 100% yield loss in susceptible cultivars. However, there are some Minnesota cultivars that have shown resistance to stem rust, including that caused by Ug99. It is therefore important to identify the QTLs for stem rust resistance in this germplasm. Association mapping is one of the most common method used to detect QTLs and genetically characterize germplasm. Our objective was to identify QTLs for resistance to the Ug99 family of stem rust pathogen races in a collection of 384 spring wheat breeding lines from the University of Minnesota.

The germplasm was screened for stem rust both in the field and as seedlings in a greenhouse. Field screening in Kenya and Ethiopia (2016 and 2017) facilitated data collection on the germplasm response to virulent races of the Ug99 race group. The seedling screening was done at the USDA-ARS CDL BSL3 greenhouse using TTKSK and TRTTF races. The germplasm was genotyped using the wheat 90K SNP Chip. The data was then analyzed using the GAPIT package in R using the Q+K model.

Significant QTLs were detected in both field seasons in Kenya but none were detected in Ethiopia. Additionally, significant QTLs for resistance to TTKSK and TRTTF races were detected in the greenhouse. Resistance to TTKSK in the greenhouse seemed to be temperature sensitive, with different QTLs being detected at different temperatures.

P1068: Wheat, Barley, Oat, and related

Identification of a Candidate Gene for *Sr9h*-Mediated Wheat Stem Rust Resistance by MutRenSeq

Matthew N. Rouse[1], Jayaveeramuthu Nirmala[2], Burkhard Steuernagel[3], Brande Wulff[3], Miroslava Karafiatova[4], Jaroslav Dolezel[4] and Colin Hiebert[5], (1)USDA-ARS Cereal Disease Lab, St Paul, MN, (2)University of Minnesota, St. Paul, MN, (3)John Innes Centre, Norwich, United Kingdom, (4)Institute of Experimental Botany, Olomouc, Czech Republic, (5)Agriculture and Agri-Food Canada, Morden, MB, Canada

The wheat stem rust resistance gene *Sr9h* confers major-effect resistance to stem rust pathogen race TTKSK (Ug99) and maps to chromosome arm 2BL in the cultivar 'Gabo 56'. *Sr9h* is one of seven phenotypic *Sr9* alleles and the only *Sr9* allele effective to TTKSK. We report here the identification of a candidate *Sr9h* gene, by the rapid MutRenSeq approach. A total of 1603 EMS-mutagenized M_2 families were screened with race TTKSK. We identified eight TTKSK-susceptible mutants that shared greater than 99% genome identity with Gabo 56 based on the 90K SNP chip. Nonsense or missense mutations were identified in the same NB-LRR candidate gene for seven of the eight mutants. A KASP marker derived from the candidate NB-LRR gene co-segregated with TTKSK resistance in two populations but appears to be a part of an NB-LRR gene family with multiple copies and pseudo genes based on the syntenic 2BL region of Chinese Spring and wild emmer wheat. Therefore we are currently sequencing chromosome 2B from Gabo56 and CM664, a second line with *Sr9h*, to fully characterize the *Sr9h* locus. Cloning the *Sr9h* gene and understanding the variation and unique phenotypic diversity underlying this complex locus will greatly enhance our understanding of the molecular mechanisms of resistance and race specificity and could provide extensive knowledge for long-term projects including the development of new resistance alleles and for the deployment of durable resistance.

P1069: Wheat, Barley, Oat, and related

Visualizing the Wheat Genome in Ensembl Plants

Guy Naamati, EMBL-EBI, Hinxton, United Kingdom

The Ensembl Plants project offers an integrative collection of interfaces for accessing and comparing genome-scale data for 53 plant species (Jan 2018). The Ensembl Genome Browser allows visualization and analysis of plant genomic sequences including gene annotation, genetic variation and comparative genomics. *Triticum aestivum* (bread wheat) is a major global cereal grain essential to human nutrition with a challenging genome size. In recent years there have been tremendous advancements in wheat genome assemblies and other data, and Ensembl Plants is an optimal platform for visualization and analysis of this important and interesting genome.

P1070: Wheat, Barley, Oat, and related

Draft Genome Sequence of Karnal Bunt Pathogen (*Tilletia indica*) of Wheat

Soma Marla, ICAR.NBPGR, New Delhi, India

Karnal bunt of wheat is a serious quarantine disease caused by hemi-biotrophic fungus, *Tilletia indica*. Despite its economic importance, little knowledge is known about molecular pathogenesis including various pathogenic determinants and avirulnce factors. *T. indica* genome is complex as reported genes do not share high degree of homology with other closest basidiomycete fungi. *T. indica* genome was sequenced employing hybrid approach of PacBio Single Molecule Real Time (SMRT) and Illumina HiSEQ 2000 sequencing platforms. The genome was assembled into 10,957 contigs, with a total size of 26.7 Mb. We predicted 11,535 putative genes, which were annotated employing Gene Ontology databases. Functional annotation of Karnal bunt pathogen genome, classification of identified effectors into protein families revealed interesting functions related to pathogenesis, several biological, cellular and molecular functions detected that are related to mating, zoospore development, signaling, host surface attachment, cell wall degrading enzymes, translocation and several others. Work is in progress to improve genome coverage and identification of potential effectors that could serve as molecular targets for development of diagnostic markers and new fungicide markers.

P1071: Wheat, Barley, Oat, and related

HIPP1-V from *Haynaldia Villosa* Positively Regulated Broad Spectrum Powdery Mildew Resistance in Common Wheat

Heng Zhang, Yingbo Li, Jia Zhao, Jin Xiao, Haiyan Wang and Xiue Wang, Nanjing Agricultural University, Nanjing, China

Powdery mildew, caused by *Bgt* (*Blumeria graminis* f. sp. *tritici*, *Bgt*), is one of the most serious fungal diseases of wheat. We previously cloned an E3 ligase coding gene *CMPG1-V* from *H. villosa* and proved it positively regulate powdery mildew resistance of wheat. To elucidate the resistance pathway mediated by *CMPG1-V*, *CMPG1-V* was used as a prey to screen a Y2H library. A total of 17 candidate interaction clones were identified, among which CMIN2 was annotated as a farnesylation protein, belonging to HIPP (Heavy metal associated isoprenylated plant protein) type. CMIN2 contains the full length gene, with an ORF of 456 bp, encoding a protein with 151 amino acids. The protein has a HMA domain and a farneyslation motif in the C terminal. The gene is designated *HIPP1-V*. Pull-down assay proved that HIPP1-V interacted with CMPG1-V, could be ubiquitinated but not degraded by CMPG1-V.

The expression of *HIPP1-V* was rapidly induced by *Bgt* infection in *H. villosa*. Transient overexpressing of *HIPP1-V* in a powdery mildew susceptible wheat variety Yangmai 158 prevented the haustorium formation. Stable transformationin Yangmai 158 enhanced its broad spectrum resistance to Bgt, When silencing the *HIPP-V* in powdery mildew resistant genotype *Triticum durum-H.villosa* amphiploid, the resistance was alleviated. All these indicate that *HIPP1-V* plays positive role in powdery mildew resistance.

P1072: Wheat, Barley, Oat, and related

High Level of Structural and Sequence Divergence between Homologous Regions of Bread Wheat and *T. militinae* within the Powdery Mildew Resistance Locus Qpm.Tut-4A

Eva Janáková[1], Irena Jakobson[2], Hilma Peusha[2], Michael Abrouk[1], Monika Škopová[1], Ljudmilla Timofejeva[2], Hana Simkova[3], Jan Safar[1], Jan Vrana[4], Jaroslav Doležel[1], Kadri Järve[2] and **Miroslav Valárik**[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia, (3)IEB, Olomouc, Czech Republic, (4)Experimental Botany, Olomouc, Czech Republic

Introgression of *QPm.tut-4A* locus from *Triticum militinae* into the distal end of bread wheat chromosome 4AL confers improved resistance against powdery mildew. The locus was high-density mapped and delimited to 0.024 cM using 8327 individuals and 75 markers. Using additional 2052 *ph[1]* lines seven new recombinations were identified. After chromosome walking, final flanking markers *owm169* and *owm228* were mapped and the region was found 640.8 kbp and 480.2 kbp long in cv. Chinese Spring (CS) and *T. militinae* (TM), respectively. The cM/Mb ratio is much smaller compared to these commonly found at the end of wheat chromosomes. The sequenced region was annotated and 16 and 12 protein coding genes were identified in CS and TM, respectively. Out of them, seven CS and six TM genes were not syntenic. Furthermore, intergenic regions do not show a significant similarity between CS and TM. The TM region containing the remaining six genes has a syntenic counterpart in CS, but that region was duplicated and one of the duplications was inverted. The duplication and inversion were accompanied by gene loss and four of the TM genes have their counterparts in both duplicated regions in CS. Finally, three genes from the CS region do not have their homologs in the TM region. These structural and sequence differences are major reasons for the discrepancy between the expected and observed cM/Mb ratio. This work was supported by award LO1204 from the National Program of Sustainability I and by the Estonian Ministry of Agriculture.

P1073: Wheat, Barley, Oat, and related

Detoxification of Mycotoxins as a Source of Resistance to Fusarium Head Blight: From *Brachypodium distachyon* to *Triticum aestivum*

Miriam Gatti[1], Frederic Choulet[2], Florence Cambon[3], Catherine Macadre[1], Caroline Tassy[4], Florence Guerard[1], Thierry Langin[3] and Marie Dufresne[1], (1)Institute of Plant Sciences Paris-Saclay, Orsay, France, (2)INRA GDEC, CLERMONT-FERRAND, France, (3)GDEC, INRA, Clermont-Ferrand, France, (4)INRA UMR 1095 GDEC, Clermont-Ferrand, France

Fusarium head blight (FHB) caused by fungi of the *Fusarium* genus is a widespread disease of wheat (*Triticum aestivum*) and other small-grain cereal crops. The main causal agent of FHB, *Fusarium graminearum*, can produce mycotoxins mainly belonging to type B trichothecenes, such as deoxynivalenol (DON) that can negatively affect humans, animals and plants. Several quantitative trait loci (QTLs) for resistance to FHB have been identified some of which have been correlated with efficient DON detoxification, mainly through the conjugation of DON into DON-3-*O*-glucose (D3G), a reaction catalyzed by UDP-glucosyltransferases (UGTs). Nevertheless, only few studies have conducted functional analyses to directly correlate DON glucosylation and resistance *in planta* and none were performed on wheat UGT gene(s). Our team, using the model cereal species *Brachypodium distachyon*, has recently demonstrated that the Bradi5g03300 UGT is able to confer tolerance to DON following glucosylation of DON into DON 3-*O*-glucose and is involved in the early establishment of quantitative resistance to FHB. In the present work, we transferred the functional analyses conducted on the model species *Brachypodium distachyon* to bread wheat. In a first approach the *B. distachyon Bradi5g03300* gene has been introduced through biolistic-mediated transformation in the wheat variety Apogee, susceptible to FHB. The phenotypic analyses conducted on homozygous transgenic wheat constitutively expressing the *Bradi5g03300* gene showed that they exhibit higher resistance to FHB as well as increased root tolerance to DON compared to the control line. In parallel, using a synteny approach between *B. distachyon* and bread wheat genomes we identified a wheat candidate gene orthologous to the *B. distachyon Bradi5g03300* gene. This wheat gene after validation through gene expression pattern during wheat infection, was introduced by transformation into *B. distachyon* to rapidly determine its ability to conjugate DON into D3G *in planta* and its involvement in FHB resistance. In conclusion, this project contributes to increase the knowledge concerning the functional relationship between DON glucosylation and FHB resistance in wheat and provide candidate genes to include in selection processes.

P1074: Wheat, Barley, Oat, and related

NAC Transcription Factor* and Laccase Gene: Key Players in Deciphering FHB Resistance Mechanism in Wheat *QTL-Fhb1

Nancy Soni, McGill University, Ste-Anne-de-Bellevue, QC, Canada

Fusarium head blight (FHB) is one of the most devastating and alarming diseases of wheat around the globe. In addition to causing a loss in wheat crop yield, it also reduces grain quality with mycotoxin contamination. Among 121 quantitative trait loci (QTLs) associated with FHB resistance, *QTL-Fhb1* is considered to have major resistance effects. Wheat near isogenic lines (NILs), derived from Sumai 3 and Thatcher cross, were sequenced using Illumina HiSeq technology to capture the genes localized within the fine mapped *QTL-Fhb1* region located within a 1.27cM interval. A total of 26 genes were putatively identified, of which, wheat NAC transcription factor (*TaNAC*), which is also known as a master regulator of plant secondary cell wall biosynthesis, was found polymorphic. Also, a laccase gene (*TaLAC*) which catalyzes cell wall lignification was also found polymorphic. Associated semi-comprehensive metabolomics study revealed a few important metabolites related to phenylpropanoid and flavonoid pathway with high fold change in pathogen inoculated samples. When the *TaNAC* or *TaLAC* silenced, the fungal biomass and the disease severity increased. However, no significant change in RR metabolites observed. In-silico analysis revealed secondary wall NAC binding element (SNBE) site in the promoter region of *TaLAC*, which suggest the regulation of laccase gene by NAC transcription factor, thus, unveiling the mechanism of FHB resistance associated with *QTL-Fhb1*.

P1075: Wheat, Barley, Oat, and related

Identifying Novel Genetic Sources for FHB Resistance in Ontario Wheat

Harwinder Singh Sidhu[1], Mitra Serajazari[1], Mina Kaviani[1], Curtis Pozniak[2], Matthew Hayden[3] and Alireza Navabi[1], (1)University of Guelph, Guelph, ON, Canada, (2)Crop Development Centre, University of Saskatchewan, Saskatoon, SK, Canada, (3)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia

Fusarium head blight (FHB) caused by *Fusarium graminearum* is one of the most detrimental diseases of wheat (*Triticum aestivum* L.). Reduced yield due to fusarium damaged kernels and mycotoxin contamination causes significant economic loss. Inadequate disease control strategies render breeding for FHB resistant wheat varieties as a favorable approach. The objective of this research is to identify the genomic regions associated with FHB resistance in a Canadian Winter Wheat Diversity Panel (n=450) and develop genomic selection models for FHB resistance breeding. The diversity panel was phenotyped in two FHB nurseries in Ontario, in 2017. The disease incidence ranged from 10% to 100% with an average of 65%. Disease severity ranged from 7% to 100% with an average of 25% 21 days after inoculation. The diversity panel was genotyped using the 90K Illumina iSelect chip, which provided dense coverage for all chromosomes with more than 50K markers. Phylogenetic trees, Principal Component Analysis, and STRUCTURE analysis alluded to the presence of population structure in the panel. Genome-wide association studies, following correction for population structure, identified a genomic region on chromosome 5A (698 mbp, maf 0.43) associated with FHB severity, in addition to other suggestive QTLs in multiple chromosomes. This research is expected to further the development of a source wheat germplasm as well as optimizing a genomic selection breeding strategy for FHB resistance breeding.

P1076: Wheat, Barley, Oat, and related

Identification of Quantitative Trait Loci (QTL) Associated with Fusarium Head Blight and Septoria Resistance in a Maxine/ Redeemer Winter Wheat Population

Ljiljana Tamburic-Ilincic and Silvia Barcrellos Rosa, University of Guelph, Ridgetown, ON, Canada

Fusarium head blight (FHB) and Septoria tritici blotch (STB) are important wheat diseases in North America. The objective of this study was to map loci associated with FHB traits, STB and plant height in a Maxine/Redeemer winter wheat population. Evaluation of FHB and STB resistance was performed using spray inoculation of a mixture of *F. graminearum* and *Septoria tritici blotch* isolates, respectively and under natural infections in replicated trials across three environments in Ontario, Canada. FHB disease incidence and severity were recorded and FHB index was calculated. For both diseases, the population showed a continuous distribution pattern and transgressive segregation of progeny. DArT markers were used to generate a genetic map and quantitative trait loci (QTL) analysis were performed by evaluating 105 doubled-haploid lines. FHB resistance QTL were identified on chromosome 2A, 4A, 6A, 3B, 4B, 2D and 3D, while QTL for STB were identified on chromosome 4B and 7A. Plant height QTL were identified on chromosome 4A, 6A, 4B and 2D. QTL identified in this study will be used in winter wheat breeding programs using marker assisted selection (MAS).

Key words: Fusarium head blight, Septoria tritici blotch, winter wheat, quantitative trait loci, marker assisted selection

P1077: Wheat, Barley, Oat, and related

Mapping QTL for Fusarium Head Blight Resistance in Canadian Spring Wheat AC Barrie

Dinushika Thambugala[1], Anita Brûlé-Babel[2], George Fedak[3], Maria Antonia Henriquez[1], Adam Foster[4], Richard Martin[4], Brent McCallum[1], Jeannie Gilbert[1], Barbara Blackwell[3], Dean Spaner[5], Curtis J Pozniak[6], Amidou N'Diaye[6] and Curt McCartney[1], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)University of Manitoba, Winnipeg, MB, Canada, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (4)Agriculture and Agri-Food Canada, Charlottetown, PE, Canada, (5)University of Alberta, Edmonton, AB, Canada, (6)University of Saskatchewan, Saskatoon, SK, Canada

Breeding for resistance to Fusarium head blight (FHB) in Canadian spring wheat is hampered by a poor understanding of genetics of resistance, particularly native FHB resistance. Here we dissected the genetic basis of FHB resistance in the Canadian spring wheat variety, AC Barrie which possesses an intermediate level of FHB resistance. A recombinant inbred line (RIL) population from the cross Cutler/AC Barrie and a doubled haploid (DH) population of the cross AC Barrie/Reeder were evaluated for FHB resistance in multiple field nurseries. Genotyping was performed with the Illumina Infinium 90K wheat SNP beadchip. IM and ICIM analyses identified numerous QTL controlling FHB resistance in the AC Barrie/Cutler RIL population on chromosomes 1B, 2A, 2B, 2D, 3B, 4D, 5A, and 6B and Barrie contributed most of these QTL. Major QTL for FHB resistance from AC Barrie were mapped on chromosomes 3B and 6B at the expected locations of *Fhb1* and *Fhb2*. Plant height locus *Rht-D1* was identified on 4D, and *Ppd-D1* locus was mapped on chromosome 2D. An additional FHB resistance QTL from AC Barrie mapped to the same region as a QTL from Nyubai on 3BS, near the centromere (3BSc). AC Barrie has a unique haplotype at *Fhb1*, *Fhb2*, and 3BSc relative to known resistance sources such as Sumai-3, Wuhan-1, and Nyubai. The DH population of the cross AC Barrie/Reeder is also being studied and results will be presented at the meeting. This study provides insight into the genetic basis of FHB resistance in Canadian spring wheat variety AC Barrie.

P1078: Wheat, Barley, Oat, and related

Identification of Quantitative Trait Loci (QTL) Associated with Fusarium Head Blight Resistance in a D8006W/ Superior Winter Wheat Population

Anjan Neupane, Department of Plant Science, University of Manitoba, Winnipeg, MB, Canada, Ljiljana Tamburic-Ilincic, University of Guelph, Ridgetown, ON, Canada, Anita Brule-Babel, University of Manitoba, Winnipeg, MB, Canada and Curt McCartney, Agriculture and Agri-Food Canada, Morden, MB, Canada

Fusarium head blight (FHB) caused by *Fusarium graminearum* is a major disease of wheat in North America. FHB infection reduces grain yield, affects end-use quality, and accumulates mycotoxins such as deoxynivalenol (DON) in the grain. The objective of this research was to identify QTL associated with FHB resistance. A doubled haploid soft white winter wheat population consisting of 107 lines from the cross D8006W/Superior was used. Evaluation for FHB reaction was performed using spray inoculation of a macroconidia mixture of four *F. graminearum* isolates representing two chemotypes in replicated field disease nurseries in three locations in Canada in 2016 and 2017. Disease incidence and severity were recorded 21 days post inoculation and FHB index was calculated. Percentage Fusarium damaged kernels and DON content were measured from collected grain samples. Both parental lines showed moderate reaction across all environments for FHB traits. However, the population showed transgressive segregation for FHB reaction with a wide continuous distribution. Genotyping of the population was performed using the 90K Illumina Infinium iSelect single nucleotide polymorphism array and 5194 high quality SNP were selected for analysis. Linkage mapping and QTL analysis is under processing. This experiment will be repeated in field nurseries in 2018. Significant FHB resistance QTL identified from this project will be used in winter wheat breeding programs using marker assisted selection.

Key words: Fusarium head blight, winter wheat, host resistance, quantitative trait loci, marker assisted selection

P1079: Wheat, Barley, Oat, and related

Identification of Consistent Loci for Fusarium Head Blight Resistance in Northern European Spring Wheat through Genome-Wide Association Mapping

Susanne Windju[1], Tatiana Belova[2], Jon Arne Dieseth[1], Muath Alsheikh[1] and **Morten Lillemo**[2], (1)Graminor, Ridabu, Norway, (2)Norwegian University of Life Sciences, Aas, Norway

Fusarium head blight resistance is quantitative, highly complex and divided into several different resistance types. QTL that are effective against several of the resistance types would be a valuable contribution for resistance breeding against this devastating wheat disease. A panel of 299 spring wheat lines with different geographical origin was tested in spawn-inoculated field trials and subjected to visual FHB assessment. In addition, DON level was analysed in the harvested seed. Anther extrusion (AE) was also assessed, in separate field trials. The panel was genotyped with the Affymetrix 35K SNP chip. Eight QTL, significant in three or more testing environments, were detected associated with both FHB and DON. These QTL were detected on chromosomes 1AS, 1AL, 2BL, 3B, 4AL, 5AL, 7AS and 7BS. AE was negatively correlated with FHB and DON, and association mapping could reveal seven AE QTL that coincided with the QTL detected for FHB and DON. The lines tested in the wheat panel harboured from zero to all the detected QTL, and the results show that resistance can be significantly increased by combining several of these resistance alleles. This information enhances the possibility to select crossing parents to obtain varieties more resistant to FHB and DON.

P1080: Wheat, Barley, Oat, and related

A Major Tan Spot Race-Nonspecific Resistance Gene in Tetraploid and Hexaploid Wheat

Justin D. Faris[1], Megan Overlander[1], Gayan Kariyawasam[2], Arron H. Carter[3] and Zhaohui Liu[2], (1)USDA-ARS, Fargo, ND, (2)North Dakota State University, Fargo, ND, (3)Washington State University, Pullman, WA

Tan spot, caused by the necrotrophic fungus *Pyrenophora tritici-repentis* (*Ptr*), is a major foliar disease of both common and durum wheat. Over the past few decades, research has revealed that wheat-*Ptr* interactions are based on an inverse gene-for-gene system, where pathogen-secreted necrotrophic effectors (also known as host-selective toxins) induce susceptibility when recognized by dominant sensitivity genes in the host. However, a few race-nonspecific resistance QTLs have also been reported. In 2005, Faris and Friesen reported a race-nonspecific QTL

with major effects on chromosome 3B in the Brazilian hard red spring wheat line BR34, and Kariyawasam et al. (2016) reported a QTL in the same region in the soft white spring wheat cultivar 'Penawawa'. Here, we evaluated the Langdon durum–*Triticum dicoccoides* accession Israel-A chromosome substitution lines (LDN-DIC) for reaction to all races. With the exception of LDN-DIC 3B being highly resistant, LDN and all the LDN-DIC lines were moderately to highly susceptible. A recombinant inbred chromosome line population derived from LDN x LDN-DIC 3B was used to map the location of a single dominant resistance gene using SSR markers. In addition, chromosome 3B linkage maps in the BR34- and Penawawa-derived mapping populations were reconstructed using the Illumina 90K SNP array and SSRs, and the disease data was reanalyzed. Comparative mapping indicated that BR34, Penawawa, and *T. dicoccoides* accession Israel-A all likely possess the same chromosome 3B tan spot resistance gene. Current progress on marker development and deployment of the gene will be presented.

P1081: Wheat, Barley, Oat, and related

Differential Expression of the Necrotrophic Effector Gene *SnTox1* in the Wheat Pathogen *Parastagonospora nodorum*

Amanda R. Peters[1], Justin D. Faris[2] and Timothy L. Friesen[2], (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

Septoria nodorum blotch (SNB) is a major foliar disease on wheat, and is caused by the necrotrophic fungus *Parastagonospora nodorum*. The wheat-*P. nodorum* pathosystem involves the recognition of pathogen-produced necrotrophic effectors (NEs) by corresponding wheat NE sensitivity genes. Recognition leads to necrotrophic effector-triggered susceptibility and ultimately disease. In this study, we evaluated a recombinant inbred population that segregates for the NE sensitivity genes *Snn1* and *Tsn1*, which recognize the *P. nodorum* NEs SnTox1 and SnToxA, respectively. The *P. nodorum* isolates Sn2000, which produces both SnTox1 and SnToxA, and Sn2000KO6-1, which is essentially identical to Sn2000 but harbors a disrupted *SnToxA* gene, were inoculated onto the population and disease was evaluated. For Sn2000, the *Tsn1*-SnToxA interaction explained 32.7% of the disease variation, with *Snn1*-SnTox1 explaining 7.1%. When Sn2000KO6 was inoculated onto plants, the *Tsn1*-SnToxA interaction was not significant and the *Snn1*-SnTox1 interaction explained 30.2% of the disease variation. Relative quantitative PCR experiments indicated that the level of *SnTox1* expression was significantly higher at almost every time point in Sn2000KO6 compared to Sn2000, suggesting that *SnToxA* may downregulate *SnTox1* expression in Sn2000. Differential expression of pathogen NE genes influences the overall importance of an interaction in its contribution to disease, providing another layer of complexity to this pathosystem.

P1082: Wheat, Barley, Oat, and related

Tolerating the Enemy: Breeding for Barley Yellow Dwarf Tolerance in Wheat

Paula Silva[1], Alexandria Kieffaber[1], Xu Wang[1], Richard Brown[1], Byron Evers[1], Shuangye Wu[1], Dorith Rotenberg[2], Allan K. Fritz[3] and Jesse Poland[1], (1)Department of Plant Pathology, Kansas State University, Manhattan, KS, (2)Department of Plant Pathology, North Carolina State University, Raleigh, NC, (3)Department of Agronomy, Kansas State University, Manhattan, KS

In the central great plains, barley yellow dwarf (BYD) is one of the most important viral diseases affecting wheat. Yield losses are highly variable depending on the environment, management practices, and the genetic background. The absence of wheat varieties with good levels of resistance and/or tolerance, and the difficulty to phenotype BYD symptoms makes breeding for BYD tolerance extremely challenging. Nevertheless, few breeding programs have included BYD tolerance as a major goal. In this study, we phenotyped and genotyped a bi-parental mapping population consisting of 310 lines, derived from the cross between 'Lakin' (moderately tolerant) and 'Fuller' (moderately susceptible). Phenotypic data were collected for BYD severity (0 – 100% visual scale), plant height and yield. Additionally, unmanned aerial systems with a MicaSense RedEdge camera was used to record plant height, vegetation indices, and individual spectral bands. The adjusted means for each trait were obtained using a randomized completed block design with two replications, defining all the predictors as fixed effects. Overall, heritability ranged between 0.72 and 0.83, with some traits being highly correlated. QTL mapping using ~14000 SNP markers obtained through GBS revealed a major effect QTL for BYD on chromosome 2D and small effect QTLs on chromosomes 1B, 2B and an additional QTL without a known position. Additional QTLs were identified for the other phenotyped traits, some of them positioned on the same chromosomes that the QTLs for BYD. Moreover, we were able to identify some promising wheat lines that tolerate the disease without compromising yield potential.

P1083: Wheat, Barley, Oat, and related

Tetraploid Wheat Germplasm Diversity Scan based on the Durum Wheat Genome Assembly

Marco Maccaferri[1], Anna-Maria Mastrangelo[2], Elisabetta Mazzucotelli[3], Raj K Pasam[4,5], Reem Joukhadar[4], Danara Ormanbekova[6,7], Silvio Salvi[8], Daniela Marone[9], Simona Corneti[10], Sara G. Milner[11], Shiaoman Chao[12], Steven S. Xu[12], Nicola Pecchioni[13], Matthew Hayden[14], Roberto Tuberosa[7], Luigi Cattivelli[15] and The International Durum Wheat Genome Sequencing Consortium[16], (1)DipSA, Department of Agricultural Science, University of Bologna, Bologna, Italy, (2)CREA-MAC, SS42, 24126 Bergamo, Foggia, Italy, (3)CREA-GB,, Fiorenzuola d' Arda, Italy, (4)Department of Economic Development, Jobs, Transport and Recourses, AgriBio Centre for AgriBioscience, Bundoora, VIC, 3083, Australia., Bundoora, Australia, (5)DEDJTR, Biosciences Research, AgriBio, Melbourne, Australia, (6)DipSA, University of Bologna, Bologna, Italy, (7)Department of Agricultural Sciences, University of Bologna, Bologna, Italy, (8)DipSA - University of Bologna, Bologna, Italy, (9)CREA-CI,, Foggia, Italy, (10)DipSA, University of Bologna,, Bologna, Italy, (11)IPK Gatersleben, Stadt Seeland, Germany, (12)USDA-ARS, Fargo, ND, (13)CREA - Research Centre for Cereal and Industrial Crops, Foggia, Italy, (14)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia, (15)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (16)CREA Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy

The genome of modern durum wheat (DW) cultivar *Svevo* has been assembled based on a combination of whole genome shotgun sequencing (270X), NRGene deNovoMagic assembler, high-resolution genetic mapping obtained from the cross between Svevo DW and Zavitan wild emmer wheat (WEW) and scaffold ordering based on chromosome conformation capture sequencing (Hi-C). The assembly consisted of 9.96 Gb of ordered sequences with 66,559 high-confidence (HC) genes. We used this resource to investigate the genetic diversity and ancestry of tetraploid wheat germplasm. iSelect 90K SNP array was used to genotype a global collection of 1,858 non-redundant accessions covering the

whole range of tetraploid genetic resources from WEW, cultivated emmer (CEW), durum landraces (DWL) and modern durum cultivars (DWC). We performed a whole-genome scan for population genetic structure, selective sweeps together with the tetraploid QTLome projection. Average whole-genome genetic diversity were p_{WEW}= 0.285, p_{CEW} = 0.254, p_{DWL} = 0.201, p_{DWC} = 0.192, with an overall WEW-DWC decrease in diversity equal to 32.6%. Diversity depletions were more relevant in peri-centromeric regions (p_{WEW_C} = 0.269, p_{DWC_C} = 0.151) as compared to the highly-recombinogenic distal regions (p_{WEW_R} = 0.287, p_{DWC_R} = 0.250). From WEW to DWC, 68 chromosome regions were subjected to diversity depletion, affecting up to 38% of the genome in total: 19 of these were associated to WEW-CEW transition, 41 to CEW-DWL and 8 to DWL-DWC. The gene content of these regions is being explored in relation to known QTL content and haplotype analysis. Overall, the analysis pointed out the chromosome regions subjected to strong selective sweeps during the domestication and breeding selection, on one side, and those regions that would benefit from targeted genetic diversity restoration on the other side.

P1084: Wheat, Barley, Oat, and related

Genomic Prediction and Genome-Wide Association Study of Grain Yield, Kernel Weight, and Kernel Number in a Durum Wheat Breeding Population

Evan Salsman, North Dakota State University, Fargo, ND

North Dakota leads the United States in acreage and production of durum wheat [*Triticum tugidum* L. subsp. *durum* (Desf.)]. Improvements in durum grain yield can result in substantial increases in profit for both farmers and the state. To date, all cultivar yield improvement in the North Dakota State University Durum (NDSU) Wheat Breeding Program has been achieved with phenotypic selection on replicated-plot yield trials in late generations. Applying genomic selection (GS) in earlier generations prior to replicated-plot testing could potentially increase genetic gain if the prediction model has merit. To understand the prospect of GS in the NDSU durum germplasm, unbalanced yield trials including approximately 1,000 breeding lines were used to generate GS models and predict breeding values of lines from the 2015 and 2016 generations. Generally, forward prediction accuracies increased as lines from additional breeding generations were added to the model. Forward prediction accuracies for the 2016 generation were 0.44, 0.42, and 0.35 for grain yield, kernel weight, and kernel number, respectively. Additionally, genome-wide association mapping revealed quantitative trait loci (QTL) for kernel weight. This information can further our understanding of genetic gain plant breeders can expect when applying GS to complex traits in an active wheat breeding program.

P1085: Wheat, Barley, Oat, and related

Identification of QTLs Associated with Kernel Texture Variation in a Soft-Kernel Durum Wheat (*Triticum turgidum* ssp. *durum*) Population

Maria Itria Ibba[1], Alecia Kiszonas[1], M.C. Simeone[2] and Craig F. Morris[3], (1)Washington State University, Pullman, WA, (2)University of Tuscia, Viterbo, Italy, (3)USDA-ARS Western Wheat Quality Lab, Pullman, WA

Kernel texture is one of the major determinants of wheat quality. This trait is primarily controlled by the *Puroindoline* genes, located at the *Hardness* (*Ha*) locus on the short arm of chromosome 5D. However additional factors contribute to minor variations in endosperm texture. Durum wheat (*Triticum turgidum* sbsp. *durum*) lacks the *Ha* locus and, therefore, its kernels exhibit an extremely hard texture that limits its end-uses. Recently, the *Puroindoline* genes from the chromosome 5DS of common wheat (*T. aestivum* L.) were introgressed into the durum wheat cultivar Langdon through the *Ph1b*-mediated homoelogous recombination, thus obtaining soft-textured kernel durum wheat lines. In the present study, soft durum wheat line Langdon 1-678 was crossed with the durum wheat variety Creso. The progeny were analyzed for kernel texture through the single kernel characterization system (SKCS) and only the lines exhibiting a hardness index (HI) < 40 were advanced, obtaining 590 soft-textured kernel F_6 lines. These lines were phenotyped through SKCS and exhibited a wide variation of kernel hardness (HI ranging from -0.3 to 37). In order to identify the genetic factors associated with variation of this phenotype, the same lines were genotyped using a targeted amplicon sequencing (TAS) approach. The identification of QTLs significantly associated with kernel hardness is in progress. To date, this is the first study to investigate the genetic control at the basis of endosperm texture in durum wheat. These results will facilitate the selection of soft durum wheat lines with superior milling properties and novel end-use applications.

P1086: Wheat, Barley, Oat, and related

MAGIC Yield: Using an Eight Founder Population for the Genetic Dissection of Yield and Yield Components in UK Winter Wheat

Benedetta Saccomanno, NIAB, Cambridge, United Kingdom

Multiparent advanced generation inter-cross (MAGIC) populations are a powerful mapping resource in crop genetics for the dissection of complex traits, previously hindered by relatively low genetic recombination and allelic diversity of traditional bi-parental populations. Wheat (*Triticum aestivum* L.) is a major arable crop of global importance, covering 1.6 million hectares in the UK alone (AHDB survey, 2017). Breeders and farmers must continue to improve wheat grain yield and yield stability to help meet demand from an increasing population, and to ensure food security in the face of the effects of climate change. The Magic Yield project helps address these problems by using an eight-founder MAGIC population (Mackay *et al.* 2014), consisting of 1,000 lines created by inter-crossing eight elite UK winter wheat varieties over three generations, to study the genetic basis of yield and yield components. With the participation of five wheat breeding companies, we conducted field trials at five UK sites for two consecutive years, phenotyping yield and a suite of pre- and post-harvest yield components. Phenotypic data coupled with Illumina iSelect 90k SNP genotype data (Gardner *et al.* 2016) allowed the detection of a total of 76 quantitative trait loci (QTL) across all year, trait and site combinations. Flanking markers for selected QTL were converted to Kompetitive Allele Specific PCR (KASP) markers to aid fine-mapping and consequent characterization of genes controlling yield. Ultimately, the resources generated will aid the selection of wheat lines with improved performance within breeding programs, for the downstream benefit to growers and end-users.

P1087: Wheat, Barley, Oat, and related

Comparison of Durum Wheat and Wild Emmer Genomes Provides Insights into Genomic Diversity in Tetraploid Wheat

Luigi Cattivelli, CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy and . The International Durum Wheat Genome Sequencing Consortium, CREA Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy

The domestication of wild emmer wheat ~10,000 years ago by early agrarian societies led to the selection of modern durum wheat widely grown today, mainly for pasta. We report the fully-assembled genome of a modern durum wheat variety (cv. Svevo) and present, via comparison with the previously published genome of wild emmer accession Zavitan, a genome-wide account of the modifications imposed by 10,000 years of selection and breeding. The durum wheat genome was assembled with the NR-Gene DeNovoMAGIC^{TM} pipeline (N50 = 6 Mb) and ordered by chromosome conformation capture sequencing (Hi-C), resulting in 14 pseudomolecules plus one group of unassigned scaffolds. A total of 66,559 high-confidence (HC) genes have been identified on the durum wheat assembly. This first genome-wide comparison between a wild and cultivated form of tetraploid wheat revealed several thousand copy-number and presence-absence variations with significantly expanded gene families in durum wheat (e.g. for disease resistance), as well as of widespread polymorphism with putative impacts on gene function. While the gene sets of durum wheat and wild emmer are highly similar, the compositions of the pseudogene sets differ in both number and enrichment for particular GO categories. Inspection at the pseudogenes in syntenic regions of durum wheat and wild emmer indicates potentially distinct duplication and pseudogenization dynamics. The comparison of the two genomes offers an overall picture of the genomic diversity between the cultivated tetraploid wheat and its wild relative progenitor.

P1088: Wheat, Barley, Oat, and related

Mutations in the Branched Head Homoeo-Allele *Bht-B1* Modify Inflorescence Architecture in Tetraploid Wheat

Gizaw M Wolde, Leib-Inst Plant Gen & Crop Plant Res, OT Gatersleben, Stadt Seeland, Germany

Inflorescence morphology directly affects the reproductive success and yield of crops. The wheat inflorescence, also known as spike, forms an unbranched inflorescence where individual spikelets are arranged distichously on the central axis of the spike, the rachis. Previously, we reported the causative mutation in the *branched head*^{t} (*bh*^{t}) gene of tetraploid wheat (*TtBH-A1*) being responsible for the loss of spikelet meristem identity, converting the non-branching wheat spike into a branched spike. Since spike-branching in wheat is a quantitatively inherited trait, we further performed whole-genome quantitative trait loci (QTL) analysis and Genome Wide Association Scans (GWAS) based on 146 recombinant inbred lines (RILs) and a collection of 302 tetraploid wheat accessions, respectively. Results showed that besides the previously found gene at the *bh*^{t}*-A1* locus on the short arm of chromosome 2A, mutations in the homoeologous gene, *TtBH-B1*, was linked to the increased penetrance and expressivity of the supernumerary spikelet (SS) and /or mini-spike formation during spike-branching thereby increasing spikelet and grain number per plant. Furthermore, we developed *bh*^{t}*-A1* Near Isogenic Lines (*bh*^{t}*-A1*-NILs) using an elite durum wheat cultivar, Floradur, for the molecular genetic dissection of the wheat spike morphogenesis and the agronomic implications of the homoeo-allele(s) for increasing grain yield production in wheat.

P1089: Wheat, Barley, Oat, and related

Genetic Dissection of Morphological and Phenological Traits Associated with Domestication Syndrome in Durum × Wild Emmer Wheat RIL Population

Andrii Fatiukha, Tzion Fahima, Valentina Klymiuk, Tamar Krugman and Abraham B. Korol, Institute of Evolution and the Department of Evolutionary and Environmental Biology, Faculty of Science and Science Education, University of Haifa, Haifa, Israel

Domestication and subsequent evolution under domestication of wheat caused substantial genetic changes, which affected plant morphology, physiology and phenology. Morphological characters, such as compactness of spikes, the number of side shoots, can be mentioned as domestication related traits in cereals. We suggest to consider the angle of side shoots (Ash) as a novel trait associated with the domestication syndrome. The objective of this study is to provide a better understanding of the antagonism between natural and man-made selection of the traits under domestication in order to identify the significant changes in phenology and morphology of wheat during domestication. We used a recombinant inbred line (RIL) population derived from a cross between *Triticum durum* (cv. Langdon) and *Triticum dicoccoides* (acc. G18-16) for mapping of quantitative trait loci (QTL) of five morphological and three phenological traits. A total of 36 QTL effects were identified that were co-located in 21 loci. Eight of these loci showed pleiotropic effects on the studied traits (including phenology). A major QTL effect of Ash, co-located with strong phenological effect, was identified on chromosome 2BL. We found that phenological loci affected the duration of flowering and development of wheat in different manners. The duration of the reproductive stage in cereals affects the development of apical meristem and many other morphological traits, such as the number of spikelets per spike and the number of side shoots. These results shed more light on shaping of wheat plant architecture and development during its evolution under domestication.

P1090: Wheat, Barley, Oat, and related

The Stems Have It: Comparative Transcriptome Analyses Reveal Expression Associated with Stem Structure in Durum Wheat

J. Allan Feurtado[1], Daiqing Huang[1], Qian Zheng[1], Parul Jain[1], Connor Dobko[2], Majid Soleimani[2], Lope Tabil[2] and Adrian J Cutler[1], (1)National Research Council Canada, Saskatoon, SK, Canada, (2)Department of Chemical and Biological Engineering, University of Saskatchewan, Saskatoon, SK, Canada

Increasing wheat productivity focuses on maximizing energy inputs into grain production while maintaining the necessary support structures such as stems, leaves, and roots for continued reproductive success. Biomechanical failure of the supporting structures has negative impacts to productivity – such is the case when shoots become displaced from an upright position and lodge as a result of stem or root anchorage failure. As part of a project focused on plant standability, we are interested to identify genes associated with differences in stem physical properties such as strength. To this end, we are characterizing a genetic cross of two durum wheats, a Canadian and an Australian cultivar, which differ in stem strength among other traits. We compared expression profiles of mRNA and small RNA (sRNA) from stem nodal and internodal regions from the parental lines and the reciprocal F1 offspring at two developmental stages. Tissue- and genotype-specific expression was identified which underlie the observed differences in stem properties. The majority of differentially expressed mRNAs and sRNAs between parents were

also expressed in F1 hybrids, demonstrating the dominant nature to which these RNAs might influence stem traits. The differential and dominantly expressed mRNA and sRNA provide a molecular genetic foundation for the optimization of stem characteristics.

P1091: Wheat, Barley, Oat, and related

Development of a Genome-Wide ChIP Derived Reference Epigenomes in Durum Wheat

Mayavan Subramani[1], Antonette Todd[1], Shaojun Xie[2], Jyothi Thimmapuram[2], Vasudevan Ayyappan[1] and Venu (Kal) Kalavacharla[1], (1)Delaware State University, Dover, DE, (2)Purdue University, West Lafayette, IN

Modifications in histones regulate gene expression by relaxing or condensing chromatin. Histone modifications play an important role in plant development and stress response. Very few studies of epigenome have been performed in wheat. The objective of this study is to serve as a blueprint for examining epigenomic regulation involved in developmental stages and stress responses in wheat using the ChIP-seq assay. The goal of our research is to produce a ChIP reference epigenome derived from tetraploid wheat. In this work, we have developed protocols for ChIP-DNA isolation from three tissues -leaf, head, and root of the durum wheat cultivar 'Langdon' to detect spatial variances in histone modification H3K4me3 (trimethylation of lysine 4 on the histone H3 protein). Strategies and statistics of read alignment will be summarized and provided as a guideline for future studies. Peak regions with H3K4me3 will be identified and characterized. Differential modification regions of H3K4me3 among different tissues will be extracted. The genes that are close to differential modification regions will be investigated to see whether these genes can explain the differences in these developmental stages. With these results, we will be able to better understand gene activation or repression in relation to this specific histone mark, paving the way for us to examine genome-wide gene regulation in wheat.

P1092: Wheat, Barley, Oat, and related

Leveraging the Tetraploid Wheat Genomes for Cloning *Cdu-B1*, a Major Gene for Cd Accumulation in Durum Wheat Grain

Sean Walkowiak[1], Neil Harris[2], Gregory Taylor[2], Krystalee Wiebe[1], P.R. MacLachlan[1], Luigi Cattivelli[3], Marco Maccaferri[4], Assaf Distelfeld[5], Andrew G. Sharpe[6], Ron E. Knox[7] and Curtis J Pozniak[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)University of Alberta, Edmonton, AB, Canada, (3)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (4)University of Bologna, Bologna, Italy, (5)Tel Aviv University, Tel Aviv, Israel, (6)Global Institute of Food Security, Saskatoon, SK, Canada, (7)Agriculture and Agri-Food Canada, Swift Current, SK, Canada

Cadmium (Cd) accumulation in the grain of durum wheat presents a serious concern for human health. As a result, durum wheat breeding programs select for low grain Cd. Differences in Cd accumulation among cultivars of durum wheat are attributed to the major-effect gene *Cdu-B1* located on chromosome 5B. The objective of this study was to identify the functional determinant of *Cdu-B1*. The fine mapped interval for *Cdu-B1* was anchored to the complete genome sequences of the durum cultivar 'Svevo' (a high Cd accumulator) and the wild emmer wheat accession 'Zavitan' (a low Cd accumulator). A sequence comparison of *Cdu-B1* between Svevo and Zavitan revealed a gene candidate, *HMA3-B1*. This gene encodes a P_{1B}-ATPase transition metal transporter and contains a 17 bp duplication in the first exon in Svevo relative to the wild-type allele in Zavitan. A molecular marker for the 17 bp duplication was used to evaluate a diverse set of breeding lines from global breeding programs and was able to identify low and high Cd accumulators with perfect precision. Furthermore, functional assays using yeast expression systems confirm a role for the wild-type *HMA3-B1* gene in regulating Cd accumulation in grain by mediating vacuolar Cd sequestration. In addition, the 17 bp duplication allele present in high Cd genotypes was non-functional. The molecular marker developed from this work is currently deployed in global breeding programs to develop wheat lines with low grain Cd.

P1093: Wheat, Barley, Oat, and related

Comparison of Durum Landraces with Northern Great Plains Adapted Cultivars and Identification of Selection Sweeps

Jason Fiedler[1], Justin Hegstad[1], Elias Elias[1], Stephen Xu[2], Shiaoman Chao[2] and Xuehui Li[1], (1)North Dakota State University, Fargo, ND, (2)USDA-ARS, Fargo, ND

Durum wheat (*T. turgidum ssp. Durum, AABB*) is a key crop for high-value food production. Modern breeding programs over the last century have developed a number of elite cultivars that are adapted for growth in the Northern Great Plains. To investigate the genomics underlying this adaptation, we compare 449 global durum lines from the Wheat Coordinated Agricultural Project with 34 advanced lines from the North Dakota State University (NDSU) Durum Breeding Program. We used genotype-by-sequencing (GBS) to identify 21,030 single nucleotide polymorphisms (SNP)s in the populations and measured genetic diversity on several scales. We find that population sub-structure designations largely agree with regional adaptation, and lines adapted to the Northern Great Plains show relatively low genetic diversity and high allelic fixation. We identified 23 genetic intervals that display differential allelic fixation between un-adapted and improved lines, suggesting that these linkage blocks are important for durum improvement. Screening potential lines for these linkage blocks could accelerate breeding efforts, and understanding the genes in these regions could shed light on the molecular characteristics of "elite" lines.

P1094: Wheat, Barley, Oat, and related

Unique Sources of Resistance to Fusarium Head Blight for Durum Wheat

George Fedak[1], Danielle Wolfe[1], Dawn Chi[1], Sylvie Cloutier[1], Allen Xue[1] and Lianquan Zhang[2], (1)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China

There was a major epidemic of FHB in the durum wheat crop in Canada in 2016. There is not as much variability for FHB resistance in the primary gene pool of *T. durum* as there is in bread wheat.

In a recent screening of synthetic hexaploids and their parents for FHB resistance by point inoculation, a number of *T. dicoccon* accessions appeared to have enhanced levels of FHB resistance. The floret infection frequencies ranged from 10-12% while the values for Langdon durum were 73%. These inoculations were repeated for a second time with similar results.

The *T. dicoccon* accessions were accessed from various gene banks. Records indicate that some of these accessions were collected in Russia and Georgia in the 1930s by N. I. Vavilov and deposited in the genebanks.

As would be expected from *T. dicoccon* accessions collected in the wild, some are deficient in useful agronomic traits. For example some are very tall and others have smaller spikes. However such traits could easily be removed by a few backcrosses so as to minimize any linked drag. On the other hand other accessions had very large seeds, a trait that could be an asset to a breeding program.
Another potential source of FHB resistance for durum wheat is that found in the amphiploid *Triticum durum* x *Hordeum chilense* with the genomes AABBHH. This source of resistance will be more difficult to integrate into durum wheat.

P1095: Wheat, Barley, Oat, and related
Population Structure of Aegilops tauschii
Le Wang[1], Naxin Huo[1,2], Karin R. Deal[1], Meiling Zou[1], Jirui Wang[3], Yong Q. Gu[2], Jan Dvorak[1] and Ming-Cheng Luo[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)USDA ARS, Western Regional Research Center, Albany, CA, (3)Triticeae Research Institute, Sichuan Agricultural University, Chengdu, Sichuan, China
Aegilops tauschii, the donor of the hexaploid wheat D genome, is distributed widely and is genetically diverse. The availability of a high-quality genomic sequence of *Ae. tauschii* enables us to analyze its population structure using whole genome resequencing strategy. Due to *Ae. tauschii*'s large genome size and abundance of repetitive sequences, whole genome exome capture was chosen to analyze the population structure of *Ae. tauschii* collections. In this study, a core collection of 63 accessions representing diversity panel of *Ae. tauschii* was subjected to capturing and sequencing with Illumina's HiSeq platform. There were 60 to 161 million high-quality reads per accession generated; of these reads, 99.4% were mapped to the *Ae. tauschii* reference genome. A total of 1,355,118 SNPs and 119,319 Indels were obtained among the 63 accessions using BWA-MEM and SAMtools. Structural analysis revealed two distinct lineages, which is in agreement with our previous observations based on the 10K Infinium assay (Wang et al. 2013), and shows that there is little gene flow between these two lineages. This work is a part of the NSF-funded Project IOS-1238231 to generate a reference genome for the genome of *Ae. tauschii*(http://aegilops.wheat.ucdavis.edu/ATGSP/).

P1096: Wheat, Barley, Oat, and related
Structural Organization and Gene Duplication in the Chromosomal Region Harboring the Alpha-Gliadin Gene Family in *Aegilops tauschii*
Yong Q. Gu[1], Naxin Huo[1], Lingli Dong[2], Shengli Zhang[3], Tingting Zhu[4], Toni Mohr[1], Susan B. Altenbach[1], Daowen Wang[2], Zhiyong Liu[2], Ming-Cheng Luo[4] and Jan Dvorak[4], (1)USDA ARS, Western Regional Research Center, Albany, CA, (2)Institute of Genetics and Developmental Biology, CAS, Beijing, China, (3)Henan Institute of Science and Technology, China, (4)Department of Plant Sciences, University of California, Davis, Davis, CA
Among the wheat prolamins important for its end-use traits, α-gliadins are the most abundant and also a major cause of food-related allergies and intolerances. Previous studies of various wheat species estimated between 25 to 150 α-gliadin genes reside in the *Gli-2* locus regions. To better understand the evolution of this complex gene family, the DNA sequence of a 1.75-Mb genomic region spanning the *Gli-2* locus was analyzed in the diploid grass, *Aegilops tauschii*, the ancestral source of D genome in hexaploid bread wheat. Comparison with orthologous regions from rice, sorghum, and *Brachypodium* revealed rapid and dynamic changes only occurring to the *Ae. tauschii Gli-2* region, including insertions of high numbers of non-syntenic genes and a high rate of tandem gene duplications, the latter of which have given rise to 12 copies of α-gliadin genes clustered within a 550-kb region. Among them, five copies have undergone pseudogenization by various mutation events. Insights into the evolutionary relationship of the duplicated α-gliadin genes were obtained from their genomic organization, transcription patterns, transposable element insertions, and phylogenetic analyses. An ancestral *GLR* gene encoding putative amino acid sensor in all four grass species has duplicated only in *Ae. tauschii* and generated three more copies that are interspersed with the α-gliadin genes. Phylogenetic inference and different gene expression patterns support functional divergence of the *Ae. tauschii GLR* copies after duplication. Our results suggest that the duplicates of α-gliadin and *GLR* genes have likely taken different evolutionary paths; conservation for the former and neofunctionalization for the latter.

P1097: Wheat, Barley, Oat, and related
Complete Chloroplast Genomes of *Aegilops tauschii* coss. and *Ae.Cylindrica* Host Sheds Light on Plasmon D Evolution
Mari Gogniashvili, Institute of Molecular Genetics, Agricultural University of Georgia, Tbilisi, Georgia and Tengiz Beridze, Agricultural University of Georgia, Tbilisi, Georgia
Hexaploid wheat (*Triticum aestivum* L., genomes AABBDD) originated in South Caucasus by allopolyploidization of the cultivated Emmer wheat *T. dicoccum* (genomes AABB) with the Caucasian *Ae. tauschii* ssp *strangulata* (genomes DD). Genetic variation of *Ae. tauschii* is an important natural resource, that is why it is of particular importance to investigate how this variation was formed during *Ae. tauschii* evolutionary history and how it is presented through the species area. The D genome is also found in tetraploid *Ae. cylindrica* Host (2*n* = 28, CCDD). The plasmon diversity that exists in *Triticum* and *Aegilops* species is of great significance for understanding the evolution of these genera. In the present investigation the complete nucleotide sequence of plasmon D (chloroplast DNA) of nine accessions of *Ae. tauschii* and two accessions of *Ae. cylindrica* are presented. Twenty-eight SNPs are characteristic for both TauL1 and TauL2 accessions of *Ae. tauschii* using TauL3 as a reference. Four SNPs are additionally observed for TauL2 lineage. The longest (27 bp) indel is located in the intergenic spacer *Rps15-ndhF* of SSC. This indel can be used for simple determination of TauL3 lineage among *Ae. tauschii* accessions. In the case of *Ae. cylindrica* additionally 7 SNPs were observed. The phylogeny tree shows that chloroplast DNA of TauL1 and TauL2 diverged from the TauL3 lineage. TauL1 lineage is relatively older then TauL2. The position of *Ae. cylindrica* accessions on *Ae. tauschii* phylogeny tree constructed on chloroplast DNA variation data is intermediate between TauL1 and TauL2. The complete nucleotide sequence of chloroplast DNA of *Ae. tauschii* and *Ae. cylindrica* allows to refine the origin and evolution of D plasmon of genus *Aegilops*.

P1098: Wheat, Barley, Oat, and related
Molecular Genetic Characterisation of Triple Rust Resistance in *Aegilops tauschii*

Naveenkumar Athiyannan[1,2], Robert A. McIntosh[3], Peng Zhang[3], Timothy Hewitt[2,3], Sutha Chandramohan[2], Sami Hoxha[3], Kerrie Forrest[4], Pippa Kay[4], Narayana Upadhyaya[2], Burkhard Steuernagel[5], Brande Wulff[5], Terese Richardson[2], Smitha Louis[2], Dhara Bhatt[2], Michael Ayliffe[2], Matthew Hayden[4], Lee Hickey[1], Yuming Long[6], Steven S. Xu[7], Evans Lagudah[2] and Sambasivam Periyannan[1,2], (1)Centre for Plant Science, Queensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane, Australia, (2)Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia, (3)Plant Breeding Institute, University of Sydney, Cobbitty, Australia, (4)School of Applied Systems Biology, AgriBio, La Trobe University, Melbourne, Australia, (5)John Innes Centre, Norwich, United Kingdom, (6)North Dakota State University, Fargo, ND, (7)USDA-ARS, Cereal Crops Research Unit, Northern Crop Science Laboratory, Fargo, ND

Bread wheat (*Triticum aestivum)* is the third most cultivated crop worldwide, and a major caloric source for the human population. Global wheat production is under constant threat due to the constant evolution of highly virulent fungal pathogens such as *Puccinia sp* that cause rust disease. Losses due to rust disease are routinely minimised through the deployment of host plant-mediated genetic resistance in commercial cultivars. However, pathogens evolve virulence to overcome this resistance. Therefore continuous supply of new sources of resistance is essential for sustainable rust management. Resistance from the wild relatives of hexaploid wheat is a valuable resource as they broaden the gene pool of available resistance genes.

In this study, CPI110672, an accession of the D genome progenitor *Aegilops tauschii,* was chosen for in-depth analysis as it resists the three wheat rust diseases namely leaf, stem and stripe rust. To characterise this triple rust resistance, we conducted genetic analysis using a mapping population derived from the cross between CPI110672 and a susceptible accession CPI110717. Through rust infection screening and 90K SNP marker analysis, the chromosome position and closely linked markers were identified. Physical maps for the chromosome region carrying these rust resistance genes were generated using the new reference genome sequences of hexaploid wheat Chinese Spring IWGSC Ref Seq v1.0[1] and the diploid *Ae. tauschii* accession, AL8/78[2,3]. Comparative genomics of these reference sequences together with contigs assembled from the sequenced genome of CPI110672 facilitated the identification of candidate genes. Functional analysis will be conducted through transformation into the rust-susceptible wheat cultivar fielder.

P1099: Wheat, Barley, Oat, and related

Introgression Lines of Triticum aestivum x Aegilops tauschii: Agronomic and Nutritional Value

Abdullah A. Jaradat, USDA-ARS, Morris, MN

Eighty-five single homozygous substitution lines (SLs) of the *Aegilops tauschii* D genome in Chinese Spring (CS) hexaploid wheat (*Triticum aestivum* L.) genetic background were evaluated for agronomic, phenotypic and ionome profiles during 3 yr of field experiments. An augmented design with a repeated bread wheat check was implemented to adjust for spatial soil variability. Percent significant pairwise differences between SLs was large for kernel phynotypic traits (78.6%) and small for the ionome (28.6%). Agronomic traits (spike harvest index, and spike fertility index) and protein content displayed large (>60.0%) significant differences between SLs. Differences among SLs accounted for 63.8 and 67.7% of variation in macro- and micronutrient concentration in kernels, respectively; and caused significant differences between macro- and micronutrients as to their functional relationships in leaves and grains; the latter showed more positive responses than the former. The ionome validation variance was smallest for CS (81.0%) as compared to 83.7% (6D) to 96.0% (4D) for SLs. Plasticity can be deduced from the ionome's phenotypic (28.6%) and genotypic (27.9%) coefficients of variation, and heritability estimate (41.5%). Most SLs exhibited high genetic potential for increasing grain micronutrient concentrations, especially Fe and Zn, by 12.0 and 8.0%, respectively, above CS as a model bread wheat cultivar. A small fraction (18%) of SLs showed yield drag compared to CS; whereas, 63% had grain yield within ±5% of CS, and the remaining 19% had significantly larger grain yield than CS. These SLs are expected to broaden the wheat genetic base to improve nutritional and agronomic traits.

P1100: Wheat, Barley, Oat, and related

A Genomewide Analysis of Phenotypic Stability in Barley

Jeffrey L. Neyhart, University of Minnesota, St. Paul, MN and Kevin P. Smith, University of Minnesota, Saint Paul, MN

Genotype-by-environment interactions (GEI) are a common challenge for plant breeders dealing with quantitative traits. Such interaction complicate efforts to develop broadly superior genotypes, but may be exploited to target genotypes for specific environments. A proxy for estimating GEI is examining the phenotypic stability (i.e. plasticity) of genotypes or genes across different environmental conditions, and knowledge of the genetic architecture of phenotypic plasticity could help inform breeding decisions when GEI are present. Our objectives were to 1) examine the genetic architecture of genotypic means and stability for complex traits in barley (*Hordeum vulgare* L.), 2) identify quantitative trait loci (QTL) influencing phenotypic stability, and 3) identify highly stable or sensitive QTL. We evaluated a diverse panel of barley breeding lines for flowering time, plant height, and grain yield across 25+ diverse environments in a highly balanced experiment. We conducted genomewide analyses to examine the genetic architecture of the genotypic mean and stability for each trait, along with the stability of individual markers. We identified significant marker-trait associations for all traits, but found little to no overlap between loci associated with the genotypic mean and those associated with linear stability, suggesting separate genetic control. Furthermore, there was moderate to high overlap between loci associated with phenotypic stability and markers exhibiting high plasticity, indicating that QTL plasticity across environments is a major contributor to phenotypic stability. Our analysis complements previous work and provides further evidence that breeding may simultaneously improve the genotypic mean and the stability of complex traits.

P1101: Wheat, Barley, Oat, and related

Developing Bioinformatics Resources to Study Meiosis in the Barley Cultivar Golden Promise

Miriam Schreiber[1], Abdellah Barakate[1], Nicola Uzrek[1], Isabelle Colas[1], Dominika Lewandowska[1], Malcolm Macaulay[1], Sybille Mittmann[2], Mikel Arrieta[2], Jonathan M Wright[3], Bernardo J. Clavijo[3], Luke Ramsay[1] and Robbie Waugh[1,2], (1)The James Hutton Institute, Invergowrie, Dundee, United Kingdom, (2)The University of Dundee at JHI, Invergowrie, Dundee, United Kingdom, (3)Earlham Institute, Norwich, United Kingdom

In barley (*Hordeum vulgare*) meiotic recombination shows a non-random pattern of recombination which is skewed towards the ends of the chromosomes. This results in around 30% of the genes in the centromeric region not recombining; a big restriction for geneticists and breeders alike. We are developing genetic resources, using different techniques and approaches to identify the roles of meiotic genes and their various natural and induced alleles in recombination in the barley cultivar Golden Promise (GP). First we developed a GP genome assembly by aligning scaffolds to the Morex reference and merging them together chromosome-wise. Second, we have established an EMS (ethyl methanesulfonate) TILLING (Targeting Induced Local Lesions in Genomes) population in GP. A first analysis of the M2 population showed a mutation frequency of approximately 10 mutations per Mb (per individual) and highlighted big differences between the individual plants. The population has and is currently being used to screen for mutations in meiotic genes. As GP is one of the few barley cultivars that can be used for genetic transformation we will be able to complement any identified mutant alleles and see if it rescues any observed phenotype. Third we've build an anther/meiocyte transcriptome covering four stages from premeiotic to pollen. This extensive transcriptome dataset will represent a reference for ongoing proteomics analyses and other comparative (e.g mutant vs. WT) experiments. It will be used to identify new genes involved in meiosis and to analyse the pattern of transcription during meiosis in genetically or environmentally perturbed plants.

P1102: Wheat, Barley, Oat, and related

Multiparental Genomic Selection for Local Adaptation in Barley (*Hordeum vulgare*)

Daniel W Sweeney, Cornell University, Ithaca, NY

Genomic prediction accuracies are generally thought to be highest within biparental families. Related biparental families contain greater genetic diversity, create higher power for QTL detection, offer a larger set of segregating progeny that are useful for selection, and may enable higher prediction accuracies than single families. A training set of ~1430 two row spring barley breeding lines from seven biparental families with a common female parent were phenotyped for Fusarium head blight, Bipolaris spot blotch, leaf rust, pre-harvest sprouting, and grain protein in two locations. These traits are critical to the establishment of malting barley production in New York state. Cross-validation was used to assess prediction accuracy within full-sib families, across half-sib families, and across all families in all traits. Inclusion of significant markers from genome-wide association studies as fixed effects were tested. Prediction accuracies between spatially corrected and base models were also compared. A selection index will be used to make selections from the training population that will be used to initiate a multi-year genomic selection experiment.

P1103: Wheat, Barley, Oat, and related

Environmental Driven Adaptation of Wild Barley in the Evolution Canyon

Cong Tan[1], Fei Dai[2], Camilla Beate Hill[1], Penghao Wang[3], Brett S Chapman[4], Qisen Zhang[5], Yong Jia[1], Xiao-Qi Zhang[1], Roberto A. Barrero[6], Gaofeng Zhou[1], Yong Han[1], Wenying Zhang[7], Matthew I. Bellgard[6], Robbie Waugh[8], Dongfa Sun[9], Eviatar Nevo[10], Guoping Zhang[11] and Chengdao Li[12], (1)Western Barley Genetics Alliance, Western Australian State Agricultural Biotechnology Centre, School of Veterinary and Life Sciences, Murdoch University, Murdoch, Australia, (2)Zhejiang Key Lab of Crop Germplasm, Department of Agronomy, Zhejiang University, Hangzhou, China, (3)School of Veterinary and Life Sciences, Murdoch University, Murdoch, Australia, (4)Centre for Comparative Genomics, Murdoch University, Murdoch, Australia, (5)Australian Export Grains Innovation Centre, Soth Perth, Australia, (6)Centre for Comparative Genomics, Murdoch University, Perth, Australia, (7)Hubei Collaborative Innovation Center for Grain Industry/College of Agriculture, Yangtze University, Jingzhou, China, (8)The James Hutton Institute, Invergowrie, Dundee, United Kingdom, (9)College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, (10)Institute of Evolution, University of Haifa, Mount Carmel, Israel, (11)College of agriculture and biotechnology, HangZhou, China, (12)Western Barley Genetics Alliance, Murdoch University, Murdoch, Australia

Background: Wild barley (*Hordeum spontaneum* L.) is the ancestor of cultivated barley (*Hordeum vulgare* L.) with a wide geographic distribution across highly diverse environments. The 'Evolution Canyon' I (EC) in Israel, consists of two abutting slopes, separated by only 250 m but with drastically different *micro*climates. It is an ideal microsite model to characterize incipient sympatric speciation of wild barleys and its adaptation to environmental stresses.

Result: In this study, transcriptome sequencing was first performed with ten wild barley accessions collected from two opposing slopes of the 'Evolution Canyon'. The wild barley accessions showed dramatic genomic diversity and were grouped into two apparent clusters based on population structure and principle component analysis as well as phylogenetic analysis. Then, whole genome sequencing with high-coverage was performed with another ten barley accessions including six wild barleys from two slops of the 'Evolution Canyon', two wild barley from Tibetan Plateau and two cultivars. It was further confirmed that the divergent distance between the wild barleys from the opposing slops is bigger than that between cultivar barley and Tibetan wild barley. Genomic regions with strong environmental selection were identified between two slops. Accessions from the two slopes showed dramatically response and differential gene expression genes to the drought treatment, which is the major environmental difference between the two slopes.

Conclusion: Our results provided a comprehensive evidence indicating incipient sympatric speciation of wild barleys in the 'Evolution Canyon'. Environmental divergence is a key driver for wild barley adaptation and genomic evolution between the two slopes.

P1104: Wheat, Barley, Oat, and related

Genotypic and Phenotypic Evaluation of Preharvest Sprouting in Two and Six Row Barley

Jason Walling[1], Justin Vetch[2], Leslie Zalapa[1], Robert Stougaard[3], Jamie Sherman[2] and Mike Giroux[2], (1)USDA, ARS, MWA, Cereal Crops Research Unit, Madison, WI, Madison, WI, (2)Montana State University, Bozeman, MT, (3)Montana State, Kalispell, MT

Preharvest Sprouting (PHS) is a problem negatively affecting both yields and quality of cereal crops grown world-wide. Preharvest Sprouting can be generalized as the propensity of a seed to begin germination while still on the parent plant and is most widely observed in regions with high humidity and/or excessive periods of rain. Barley with signs of PHS is rejected for malt and can only be sold as feed, results in a loss to the grower of about half the value. Preharvest sprouting is a complex trait involving contributions from both multi-genic and environmental

factors. Recently, a gene (*TaPHS1*), was described in *Triticum aestivum* (wheat) whose variable genotype and specific gene expression were associated with wheat lines that show either resistance or susceptibility to PHS. Here we present the exonic sequencing and genotypic characterization of the barley (*Hordeum vulgare*) homolog *HvPHS1* in over 120 barley lines. Additionally, we evaluated each of these lines for dormancy using standard germination tests and also for PHS by challenging intact heads to sprout in an artificial rain chamber.

P1105: Wheat, Barley, Oat, and related

Identification of Genes Deficient in Barley (*Hordeum vulgare* L.) Mutants with a Short-Culm Phenotype

Mats Hansson, Lund University, Lund, Sweden

Lodging is the process where crop plants fall over and lie on the ground due to strong winds and heavy precipitation. This problem reduces yield and increases the risk of fungal infections and pre-harvest germination. In order to avoid lodging, plant breeders utilize short-culm mutants, which often have a robust culm that can support the weight of a heavy spike. In barley (*Hordeum vulgare* L.), thousands of short-culm mutants have been isolated in breeding programs around the world. However, the current elite cultivars predominantly use loss-of-function alleles of only the *Sdw1* gene encoding gibberellin-20-oxidase. In order to reveal the genetic network underlying culm length, with the objective to provide an enlarged repertoire of genes and alleles suitable for future breeding of lodging resistant barley, we are identifying the deficient genes in short-culm barley mutants. These mutants are mainly found in the mutant groups named *brachytic* (*brh*), *semi-brachytic* (*uzu*), *erectoides* (*ert*), *breviaristatum* (*ari*), *dense spike* (*dsp*), *curly dwarf* (*cud*), *semi-dwarf* (*sdw*) and *slender dwarf* (*sld*). So far we have identified mutated genes related to brassinosteroid signalling and biosynthesis. However, we have also identified mutations in genes encoding different subunits of a heterotrimeric G protein and a leucine-rich repeat receptor-like kinase; both of unknown molecular function.

References

Braumann et al. (2017) Hereditas 155: 10
Dockter et al. (2014) Plant Physiol 166: 1912-1927
Wendt et al. (2016) PloS One 11: e0168924
Zakhrabekova et al. (2015) Plant Mol Biol 88: 609-626

P1106: Wheat, Barley, Oat, and related

Genomics of Barley Tiller Development, Many Branches on a Theme

Allison M. Haaning[1], Kevin P. Smith[1], Gina Brown-Guedira[2], Shiaoman Chao[3], Priyanka Tayagi[4] and Gary J. Muehlbauer[5], (1)University of Minnesota, Saint Paul, MN, (2)CropScience NCSU, Raleigh, NC, (3)USDA-ARS, Fargo, ND, (4)CropScience NCSU, Raleigh, NC, Raleigh, NC, (5)University of Minnesota, St. Paul, MN

Shoot architecture of barley is largely defined by the number and vigor of tillers, and the majority of grain harvested comes from tillers. Despite this, genetic regulation and other sources of variation that impact tillering throughout development are not well characterized. The main goals of this work were to (1) identify primary sources of variation that impact tiller number and rate of outgrowth and (2) to identify genes and regulatory networks important for early tiller development. For the first goal 768 genetically diverse lines, split equally between two- and six-row spike morphology, from the USDA National Small Grains Collection were genotyped by GBS and 50K SNP array and grown in the field in 2014 and 2015; and data for the following traits was collected: tiller number from two to seven weeks past-emergence, productive tiller number, plant height, days to heading, seeds per spike, fifty seed weight, stem diameter, and leaf width. Results of genome-wide association mapping and phenotypic analyses suggest tiller number and rate of development are primarily influenced by environment, spike row-type, and days to heading. For the second goal, tissue from Bowman and Morex seedlings was harvested by laser microdissection from shoot apical meristems (SAM) and axillary meristems (AXM), which form the main stem and tillers, respectively. RNA sequencing analysis identified 102 genes that were differentially expressed between SAM and AXM in Bowman and Morex, and annotation and GO term enrichment of these genes suggests that early tiller development is mediated by ethylene signaling similar to submergence response in rice.

P1107: Wheat, Barley, Oat, and related

Association Mapping of Waterlogging Tolerance in Two-Row Barley (*Hordeum vulgare* L.)

Ana Borrego-Benjumea[1], William G. Legge[1], Min Zhu[2], Meixue Zhou[3], Adam Carter[1], Jatinder Shanga[4], James R. Tucker[1] and Ana Badea[1], (1)Agriculture and Agri-Food Canada, Brandon Research and Development Centre, Brandon, MB, Canada, (2)Yangzhou University, Yangzhou, Jiangsu, China, (3)Tasmanian Institute of Agriculture and School of Land and Food, University of Tasmania, Prospect, TAS, Australia, (4)Agriculture and Agri-Food Canada, Swift Current Research and Development Centre, Swift Current, SK, Canada

Waterlogging is a major abiotic stress causing barley yield losses globally. In Canada, excessive moisture has been identified as a major issue for crops in western regions. The use of quantitative trait loci (QTL) for waterlogging tolerance in marker-assisted selection in barley breeding will accelerate the introgression of waterlogging tolerance in the Canadian germplasm. In this study, 272 two-row spring barley genotypes originated from across the world was grown at Brandon, Manitoba, Canada for two years (2016 and 2017), in order to identify QTLs for various waterlogging tolerance traits by an association mapping approach using a 50K SNP array. Abundant phenotypic variation was observed in each trait: biomass, spikes per plant, grains per plant, kernel weight, plant height and chlorophyll content. Preliminary Genome-wide association analysis was conducted for the first-year field experiment, using methods to account for population structure and minimize false-positive associations. The set of 272 two-row barley genotypes consisted of 3 clusters, based on population structure analysis. A mixed model, using both population structure and a kinship matrix (Q+K), identified approximately 100 significant SNP marker-trait associations ($P<0.001$). A total of 22 genomic regions were associated with waterlogging tolerance, with most of the putative QTLs localized on chromosomes 1H, 2H and 5H. The 22 detected putative QTLs accounted for 6.8% to 14.6% of the phenotypic variation. Putative QTLs associated with spikes per plant and plant height were identified in the same region of chromosome 2H. These field studies are complemented by seedling stage screening for waterlogging with growth chambers.

P1108: Wheat, Barley, Oat, and related

Identifying Useful Alleles for Crop Improvement: Applying State of the Art Genomic Tools, Methods and Approaches to Characterise the WHEALBI Barley Genetic Resource

Daniela Bustos-Korts[1], **Alessandro Tondelli**[2], Joanne Russell[3], Ian Dawson[3], Noemi Trabanco[4], Davide Guerra[2], Stefano Delbono[2], Stylianos Kyriakidis[3], Allan Booth[3], Chiara Ferrandi[4], Francesco Strozzi[4], Ezequiel L. Nicolazzi[4], Hakan Ozkan[5], Marta Molnar-Lang[6], Mária Megyeri[6], Mikó Péter[6], Benjamin Kilian[7], Nils Stein[8], Laura Rossini[4], Robbie Waugh[3], Luigi Cattivelli[2] and Fred A. van Eeuwijk[1], (1)Wageningen University & Research - Biometris, Wageningen, Netherlands, (2)CREA - Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda, Italy, (3)The James Hutton Institute, Invergowrie, Dundee, United Kingdom, (4)Parco Tecnologico Padano, Lodi, Italy, (5)University of Cukurova, Adana, Turkey, (6)Agricultural Institute, MTA ATK, Martonvásár, Hungary, (7)The Global Crop Diversity Trust (GDCT), Bonn, Germany, (8)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany

The EU-funded WHEALBI project (Wheat and barley Legacy for Breeding Improvement; http://www.whealbi.eu) is taking a multidisciplinary approach to identify, understand and utilise the genetic diversity available in wheat and barley cultivars, landraces and wild relatives. This is strategic to meet the challenge of increasing crop yields while reducing environmental impact. Here we focus on a carefully selected set of 400 barley accessions from extensive *ex situ* collections which cover the geographical and agro-ecological adaptive range of barley. Agronomic and life history traits, collected from multi-environment common gardens experiments across Europe, provide a unique dataset to decipher the genetic basis of adaptation to environmental conditions. A comprehensive molecular variant analysis by exome sequencing identified 1.75 million SNPs that have been used to investigate allelic variation at candidate genes driving phenotypic differences for heading date, plant height, grain weight, and awn length. We were able to relate geographical origins using site information to genotypic diversity, providing valuable information to identify novel 'adapted' alleles for future breeding under a changing climate.

P1109: Wheat, Barley, Oat, and related

Evaluation of (1,3;1,4)-β-D-Glucan in Barley (*Hordeum vulgare*) Endosperm.

Ivonne Benitez[1], Jamie Sherman[2] and **Christopher Botanga**[1], (1)Chicago State University, Chicago, IL, (2)Montana State University, Bozeman, MT

Barley is an important cereal crop with global significance, and has recently attracted several areas of research interests. Worldwide usage of barley is concentrated in the alcoholic beverage industry followed by livestock feed, and only a small amount is used for human consumption. However, because barley has a high content of the polysaccharide beta glucan it is increasingly sought after for the numerous health benefits that may result from its consumption. In contrast, barley with high beta glucan content is undesirable by the malt and alcohol industries due to its negative impact on viscosity during the brewing process. Previous studies have identified QTL loci that are associated with the biosynthesis of beta glucan located on chromosomes 1H, 2H, 5H, and 7H. In this study, we are evaluating some genotypes of barley in order to characterize them as high or low in endosperm beta glucan content, and further carry out sequence homology analysis in order to determine if there are any differences in the sequences of the beta glucan genes. We are particularly interested in identifying polymorphisms arising from SNPs and INDELS that may serve as valuable markers for use in breeding programs aimed at altering the content of beta glucan. *Funding: Funding support: USDA-NIFA Award No. 2016-70003-24775, USDA-NIFA Triticeae Coordinated Agricultural Project (T-CAP) and also by the U.S. Department of Education Award Number P382A110049.*

P1110: Wheat, Barley, Oat, and related

Impact of Colchicine Treatment on Genetic and Epigenetic Variation in Barley Regenerants.

Katarzyna Makowska, Plant Breeding and Acclimatization Institute - National Research Institute, Błonie, Poland and Joanna Machczyńska, Janusz Zimny, Piotr Bednarek, Renata Orłowska

Androgenesis is a process of induced unusual microspore development - into embryos that can grow into plants. Along the way, spontaneous doubling of the chromosome (SCD) number may occur leading to doubled haploids (DHs). Homozygosity makes DHs interesting objects for genetic and genomic research, as well as valuable material in plant breeding. In barley, SCD may range from 55% to 90% depending on genotypes. Variable levels of SCD combined with low effectiveness of androgenesis dictate an antimitotic agents application to produce DHs. Colchicine is commonly used to induce chromosome doubling (ICD) but it has negative effects such as mixoploidy, abnormal morphology and changes in gene expression that may be due to DNA methylation changes. Here we address the effect of colchicine on DNA sequence changes and alteration of the DNA methylation patterns among androgenic regenerants.

Haploid (H) green androgenic regenerants verified by flow cytometry were treated with colchicine in two concentrations (0.06% and 0.14%). The seed set of ICD plants ranged from sterile to full fertile. Control plants (without colchicine treatment), SCD regenerants and two groups of treated plants were studied using metAFLP and HPLC-RP. The global cytosine methylation was at the same level in the control and the colchicine treated samples, being at the same time higher than SCD plants. MetAFLP was performed to assess the level of sequence variation, de novo methylation and demethylation using two pairs of restriction enzymes (Acc65I/MseI, KpnI/MseI). The MetAFLP results showed differences between SCD and ICD plant what would be discussed.

P1111: Wheat, Barley, Oat, and related

Differential Expression Profiling of Barley Microspores during the Early Stages of Isolated Microspore Culture and Embryogenesis

Sébastien Bélanger, Laval University, Québec, QC, Canada, Suzanne Marchand, Université Laval, Québec, QC, Canada, Pierre-Étienne Jacques, Sherbrooke University, Sherbrooke, QC, Canada, Blake Meyers, Donald Danforth Plant Science Center, St. Louis, MO and Francois Belzile, Laval University, Quebec, QC, Canada

In barley, it is possible to induce embryogenesis in the immature microspore (male gametophyte) to obtain a diploid plant that is perfectly homozygous (doubled haploid or DH). To change developmental fates in this fashion, microspores need to engage in cellular de-differentiation, interrupting the transcriptional and translational activities leading to pollen formation, and restore totipotency prior to engaging in

embryogenesis. In this work, our objective was to characterize the transcriptome of immature barley microspores prior to (day 0) and immediately after (days 2 and 5) the application of a stress pretreatment that induces embryogenesis. A deep RNA-seq analysis revealed that microspores at these three time points exhibit a transcriptome of ~14k transcripts, ~90% of which were shared. Despite this extensive overlap, the three transcriptomes proved highly reproducible and distinct, and these differences were due to differential expression of a small set of genes (<500). The microspore response to the pretreatment applied was marked by an increased expression level of numerous Glutathione S-transferase (*GST*) and Heat shock protein (*HSP*) genes known to be involved in protection against stresses. The transition from microspore to developing embryo was marked by the induction of transcription factor genes known to play important roles in early embryogenesis and many genes involved in the synthesis and response to growth regulators. This work sheds light on the transcriptional changes that accompany an important developmental shift and provides candidate biomarkers for embryogenesis in barley.

P1112: Wheat, Barley, Oat, and related
Cloning of the Zero-Rowed Spike 1 in Barley
Shun Sakuma, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, Martin Mascher, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany, Takao Komatsuda, Institute of Crop Science, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan and Thorsten Schnurbusch, Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Stadt Seeland, Germany
Inflorescence architecture is a major determinant of the components of final grain yield in the cereals. The inflorescence can take the form of a panicle (rice, sorghum, and maize) or a spike (wheat, barley and rye). Barley's spike is composed of three spikelets (one central spikelet and two lateral spikelets) per rachis node that is a unique character of *Hordeum* species among Triticeae. Cultivated barley (*Hordeum vulgare* ssp. *vulgare* L.) produces either two-rowed (central spikelet fertile; lateral spikelets sterile) or six-rowed (complete fertility of the three spikelets) spikes. The six-rowed spike or lateral spikelet fertility is under the control of *Six-rowed spike 1* (*vrs1)*, *vrs2*, *vrs3, vrs4* and *Intermedium spike-c* (*int-c*). However, the genetic basis of three-spikelet structure in a distichous manner was not fully elucidated yet. To address this, we identified the *zero-rowed spike 1* (*zrs1*) mutant derived from mutagenesis of wild barley (*Hordeum vulgare* ssp. *spontaneum* L.). The *zrs1* mutant shows severe spikelet initiation defects and its distichous pattern is lost. At the vegetative growth the phylotaxis is normal as wildtype, however, after reproductive stage some tillers show onion-like leaf structure. We conducted genetic analysis using whole genome sequencing and RNA sequencing approach to reveal the genetic basis of the *zrs1* mutant.

P1113: Wheat, Barley, Oat, and related
Genome-Wide Association Study for Seven Traits in Spring Barley from Eastern Canada
Amina Abed, Laval University, Québec, QC, Canada and Francois Belzile, Laval University, Quebec, QC, Canada
Genome-wide association studies (GWAS) are a powerful tool to identify quantitative trait loci (QTL) determining complex agronomic traits by relating genotypes at large numbers of markers to observed phenotypes. The detection of a credible genotype-phenotype association requires accurate genotypic data but also equally reliable phenotypic information. For many complex traits subject to environmental influences, extensive data obtained in registration trials (RTs) can help meet this goal. In this study, we explored the genetic basis of seven important traits in spring barley: deoxynivalenol (DON) content in kernels, heading time (HTM), days to maturity (MAT), thousand-kernel weight (TKW), specific weight (SPW), grain yield (GYD) and plant height (PHT). A population of 200 advanced lines representing the genetic diversity of an Eastern Canadian barley breeding program were used. Phenotypic data were recovered from RTs carried out in 14 different locations from 2004 to 2014, for a total of ~20 environments (location x year). Genotypic characterization was carried out using a genotyping-by-sequencing (GBS) approach, resulting in ~40.000 polymorphic and high-quality SNP markers adequately covering all chromosomes. Population structure was examined by various methods (*P* matrix, *K* matrix and *Q* matrix) and different statistical models were tested to control it and to calculate *P*-values for marker-trait associations. We report some significant marker-trait associations obtained with the best approach. Our results demonstrated that such historical phenotypic data could be extremely valuable for performing GWAS for various quantitative traits. The significant associations obtained should provide information on trait architecture and tools for designing the next breeding program.

P1114: Wheat, Barley, Oat, and related
Novel Approaches to Analysing Ribosomal RNA Loci in Barley
Helena Toegelova[1], Martin Mascher[2], Jiri Macas[3], Tomas Beseda[1], Zuzana Tulpova[1], Jaakko Tanskanen[4], Alan H. Schulman[4], Nils Stein[2], Jaroslav Doležel[1] and **Hana Simkova**[5], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany, (3)Biology Centre CAS, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic, (4)LUKE & University of Helsinki, Helsinki, Finland, (5)IEB, Olomouc, Czech Republic
Ribosomal RNA loci pose an indispensable part of both prokaryotic and eukaryotic genomes. In eukaryotes, they are associated with nucleolus, the largest functional domain of the nucleus and the site of ribosomal biogenesis. Ribosomal DNAs (rDNAs) are mostly organized as long head-to-tail tandem arrays spanning several hundred kilobases to multiple megabases, which precludes their complete assembling from NGS data and impedes characterization of particular loci. In our study, we identified and analysed rDNA loci in barley genome by Bionano genome (BNG) mapping, a technology that visualizes short sequence motives along DNA molecules of 150 kb to 1 Mb. 45S rDNAs appear as labelled simple or compound repeats in BNG maps created using Nt.*Bsp*QI nicking enzyme. rDNA units in barley (*Hordeum vulgare*) genome maps had a size of ~9 kb, and were detected both as simple and two-label compound repeats. Only the compound variant, featured by additional *Bsp*QI site in IGS region, was identified in genome maps of wild barley (*H. spontaneum*). Combination of BNG mapping with other approaches such as Hi-C on barley mitotic chromosomes and long-read OxfordNanopore sequencing facilitated positioning and delimiting major as well as minor 45S rDNA loci in barley genome, thus building the basis for their detailed characterization.

P1115: Wheat, Barley, Oat, and related
Regulation of Barley and Powdery Mildew Interaction by Small RNAs

Qian-Hua Shen, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China
During plant-pathogen interactions, some plant and pathogen derived small RNAs have also shown to participate in trans-kingdom targeting and gene silencing. Here, we aim to identify and characterize sRNAs that may fulfill trans-kingdom functions in barley (*Hv*) and barley powdery mildew (*Bgh*) interactions. We have performed deep sRNA sequencing of 14 libraries during barley-powdery mildew interactions and identified 516 Huv-miRNAs (85 novel, 431 conserved) from barley and 84 Bgh-milRNAs (83 novel, 1 conserved) from *Bgh*. On the basis of the microRNA's cross kingdom targets we screened 25 Huv-miRNAs (10 novel) and 9 Bgh-milRNAs (8 novel) as candidate trans-kingdom miRNAs. The differential expression study on the candidate trans-kingdom miRNAs classified miRNAs in two groups; miRNAs from first group have higher expression in healthy epidermis (EPH) but expression level reached to zero in infected epidermis (EPI1) then increases with the disease progression till conidia formation. miRNAs from second group showed opposite tendency and the expression was almost zero in EPH and have higher expression in EPI1 then decreased with disease progression and reached to zero in conidia.
We performed *Agrobacterium*-mediated transient co-expression assays in *Nicotiana benthamiana* to verify miRNA-target relations. Western blot analysis showed that Hvu-miR07, Hvu-miR535 and Hvu-miR6180 can reduce the protein level of *Bgh*-BUB3 (CCU74706), *Bgh*-hypothetical protein (CCU75788) and *Bgh*-DIL protein (CCU75480) respectively. *Bgh*-milRNA38 target one barley receptor-like kinase gene (HORVU2Hr1G087540). Further experiments will be performed to test the function of candidate trans-kingdom miRNAs and their targets in barley-powdery mildew interactions.

P1116: Wheat, Barley, Oat, and related
Zinc Oxide Nanoparticles Induce Transcriptional Reprogramming that Compromises Resistance to *Pyrenophora teres f. teres* in Barley
Priyanka Deka[1], Roshan Sharma Poudel[2], Chrysafis Vogiatzis[3], Robert S. Brueggeman[2], Achintya Bezbaruah[1] and NDSU Engineering and Plant Pathology, (1)North Dakota State University, Fargo, ND, (2)Department of Plant Pathology, North Dakota State University, Fargo, ND, (3)North Carolina A&T State University, Greensboro, NC
Barley line CI5791 has effective and broad resistance to diverse isolates of the necrotrophic fungal pathogen *Pyrenophora teres f. teres* (*Ptt*), the causal agent of the disease net form net blotch (NFNB). We showed that ZnO engineered nanoparticle (NP) exposure compromises CI5791 NFNB resistance. Thus, this pathosystem can be used to characterize the effects of NPs on resistance mechanisms. A time course RNAseq analyses from leaf tissue at 0, 6, 24 and 48 hours post NP application (hpa) or pathogen inoculation (hpi) was conducted on CI5791 post ZnO NP exposure, post *Ptt* inoculation, and post dual application/inoculation with NPs + *Ptt*. The analyses identified differentially expressed genes (DEGs) in response to the treatments showing rapid responses to ZnO NPs (6 hpa) that quickly returned to basal levels (12 hpa). However, treatment with the pathogen alone and pathogen+ZnO NP, showed DEG profiles that persisted to 48hpi. The number of DEGs in the dual application/inoculation was the highest across all time-points compared to the pathogen and ZnO NPs alone. Gene ontology analysis of the DEGs revealed that the salicylic acid (SA)-signaling pathway persisted 48hpi in the dual application that resulted in compromised resistance. The SA and jasmonic acid (JA) responses returned to basal levels in the *Ptt* inoculated treatments (24hpi/hpa), which resulted in a resistance reaction. Thus, we hypothesize that the NP application resulted in extended SA responses when faced with the necrotrophic pathogen, which lead to the suppression of JA mediated necrotrophic pathogen resistance responses resulting in compromised immunity.

P1117: Wheat, Barley, Oat, and related
Fine-Mapping the *Rha4* Region – A Search to Identify a Cereal Cyst Nematode Resistance Gene in Barley
Kelvin Khoo[1], Caitlin Thompson[1], John Lewis[2], Ken Chalmers[1] and Diane Mather[1], (1)The University of Adelaide, Adelaide, Australia, (2)South Australian Research and Development Institute (SARDI), Adelaide, Australia
The cereal cyst nematode (CCN) *Heterodera avenae* Woll. is a soil-borne parasite that infects the roots of many cereal crops, leading to significant yield losses. Female nematodes become sedentary and form egg-filled cysts on the roots of successfully infected host plants. The cysts remain in the soil after harvest and rupture when conditions are optimal, allowing the eggs to hatch and repeat the infection process. The *Rha4* locus on chromosome 5H confers strong resistance against the Ha13 pathotype of *H. avenae*. Barley varieties with *Rha4* resistance have been widely deployed in Australia. To identify DNA polymorphisms linked with *Rha4*, we applied genotyping-by-sequencing (GBS) to resistant and susceptible progeny of several crosses that segregate for *Rha4*. Conversion of SNPs identified in the GBS tags to uniplex KASP™ assays yielded closely-linked SNP markers. Some of these have been applied in barley breeding and in an ongoing effort to isolate the causal gene. With genotyping and phenotyping of progeny in a backcrossing scheme with Galleon (resistant, *Rha4*) as the donor parent and Sloop (susceptible) as the recurrent parent, the region of interest has been narrowed to just 15,000 bp in the current 5H pseudomolecule of the ISBC reference genome. Further research is ongoing to resolve some minor discrepancies between genetic and physical map orders and to investigate candidate genes.

P1118: Wheat, Barley, Oat, and related
A Reference Quality Assembly and Annotation of the *Avena atlantica* Genome
Rebekah Lee[1], **Peter J. Maughan**[1], Tim Langdon[2], Jessica Schlueter[3] and Rick Jellen[1], (1)Brigham Young University, Provo, UT, (2)IBERS, Aberystwyth University, Aberystwyth, United Kingdom, (3)University of North Carolina at Charlotte, Charlotte, NC
Common oat (*Avena*) has held a significant place within the global crop community for centuries. Although its cultivation has decreased over the past century, its nutritional benefits has garnered renewed interest for human consumption. Until now there has not been a published reference genome for any of the three oat sub genomes. Here we report a quality sequence assembly, annotation and hybrid optical map assembly of the A-genome diploid *Avena atlantica*. The hybrid assembly is composed of 3417 scaffolds, spanning ~3.7 Gb, with an N50 of 11.86 Mb and a BUSCO estimated completeness of 97.6%. It is hoped that this genome sequence can escalate research within the oat community.

P1119: Wheat, Barley, Oat, and related
The Hexaploid Oat Genome

Nick Sirijovski, Leif Bulow and Olof Olsson, Lund University, Lund, Sweden
Relative to other cereals such as rice, barley and wheat, very little is know about the genetics of oat. Cultivated oat (*Avena sativa*) is a hexaploid comprised of three diploid genomes (AACCDD). It has a 1C genome of 21 chromosomes with a total size estimated to 13Gb. The large genome size and polyploidy has meant that deciphering the genetics of cultivated oat has lagged behind other cereals. Recently, oat has received much attention due to well documented health benefits of consuming this 'super food', which in turn has lead to increased production of oat-based novel foods and ingredients e.g. dairy alternatives, beta-glucan extracts, and even meat substitutes. With the fast paced development of next generation sequencing technologies, it has now become possible and affordable to undertake genome sequencing of hexaploid oat using short read technology. Herein we report on the status of the Swedish oat genome sequencing project, which is part of the newly inaugurated ScanOats research center in Lund, Sweden.

P1120: Wheat, Barley, Oat, and related
Assembly and Annotation of the Hexaploid Oat Genome
Rachel N. Walstead[1], Adam Whaley[1], Robert Reid[1], Veronica Vallejo[2], Cory Brouwer[1] and Jessica Schlueter[1], (1)University of North Carolina at Charlotte, Charlotte, NC, (2)PepsiCo, Hawthorne, NY
The hexaploid oat (*Avena sativa* L) is a staple cereal crop, used for both human consumption and animal feed. Genomic resources for oat, despite of the importance of cereals, are lagging behind many other crops. The estimated genome size of *A. sativa* is about 13GB. We aim to fully sequence, assemble, and annotate the hexaploid oat genome (2n = 6x = 42). To this aim, we have utilized PacBio RSII technologies and sequenced approximately 580 SMRT cells, achieving a coverage of approximately 40X. Continuing sequencing efforts utilizing PacBio Sequel technologies are underway, so far totaling 49 SMRT cells. We are currently assembling the genome using the Canu assembler. A draft assembly was completed with a size of ~7.5 Gb.
Annotation efforts are underway on the draft assembly, utilizing RNAseq and Iso-Seq data. This data will be integrated with predictive gene models and comparative annotations from other grass genomes using the BRAKER1 annotation pipeline (http://exon.gatech.edu/braker1.html). The genomic information obtained by this project will be a valuable resource for crop scientists and breeders.

P1121: Wheat, Barley, Oat, and related
Plans for a Hexaploid Oat Genome, and Beyond
Nick Sirijovski and Olof Olsson, Lund University, Lund, Sweden
Relative to other cereals such as rice, barley and wheat, very little is know about the genetics of oat. Cultivated oat (*Avena sativa*) is a hexaploid comprised of three diploid genomes (AACCDD). It has a 1C genome of 21 chromosomes with a total size estimated to 13Gb. The large genome size and polyploidy has meant that deciphering the genetics of cultivated oat has lagged behind other cereals. Recently, oat has received much attention due to well documented health benefits of consuming this 'super food', which in turn has lead to increased production of oat-based novel foods and ingredients e.g. dairy alternatives, beta-glucan extracts, and even meat substitutes. With the fast paced development of next generation sequencing technologies, it has now become possible and affordable to undertake genome sequencing of hexaploid oat using short read technology. Herein we report on the status of the Swedish oat genome sequencing project, which is part of the newly inaugurated ScanOats research center in Lund, Sweden. In addition, we present our plans and preliminary data for non-destructive analysis of oat seeds using neutrons, X-rays and NMR.

P1122: Wheat, Barley, Oat, and related
Mapping of the Oat Crown Rust Resistance Gene *Pc45* and its Relationship to *PcKM*
Aida Z. Kebede[1], Jayelle Friesen[2], Jim Menzies[1], Jennifer Mitchell Fetch[3], Edyta Paczos-Grzęda[4], Aaron D. Beattie[5] and Curt McCartney[1], (1)Agriculture and Agri-Food Canada, Morden, MB, Canada, (2)University of Manitoba, Winnipeg, MB, Canada, (3)Agriculture and Agri-Food Canada, Brandon, MB, Canada, (4)Institute of Plant Genetics, Breeding and Biotechnology, University of Life Sciences in Lublin, Lublin, Poland, (5)University of Saskatchewan, Saskatoon, SK, Canada
The development and effective deployment of crown rust resistant oat cultivars requires the identification, phenotypic and molecular characterization of available resistance genes. An earlier linkage mapping study identified a seedling crown rust resistance gene, temporarily designated as *PcKM* in the oat cultivars Kame and Morton. This gene was mapped to Chromosome 12D of the oat chromosome-anchored linkage map. In addition, Kompetitive Allele Specific PCR (KASP) markers closely linked to this gene were developed. It was postulated that *PcKM* was *Pc45* based upon haplotype analysis of *PcKM*-linked SNPs and comparison of the individual *P. coronata* races to which *PcKM* and *Pc45* confer resistance. To further investigate whether *PcKM* and *Pc45* are the same gene, a *Pc45* differential line was mapped in crosses involving two susceptible checks, AC Morgan and Kasztan. Seedlings of F_2 progeny and $F_{2:3}$ families generated from the crosses were inoculated with crown rust isolate *CR258* (race NTGG) in the greenhouse at the single leaf stage. Seedling ITs and KASP markers were used to create linkage maps. Results show that the IT segregation ratio for F_2 progeny (3:1) and $F_{2:3}$ families (1:2:1) of the two populations follow a single gene inheritance pattern as expected. Interestingly, the KASP markers identified to be linked with *PcKM* gene were also linked with *Pc45* gene confirming that the two genes were actually the same or very closely linked to each other.

P1123: Wheat, Barley, Oat, and related
Epigenetics of Cold Acclimation and Vernalization in *Brachypodium distachyon*
Lifeng Liu, Yuko Yoshinaga, Matthew Blow, Ronan O'Malley and John P. Vogel, DOE Joint Genome Institute, Walnut Creek, CA
Many temperate plants respond to prolonged cold exposure by increasing their freezing tolerance and transitioning to from a vegetative to a flowering state, i.e. vernalization. It is known that the histone modification of certain genes changes during vernalization. However, the role of global epigenetic changes in mediating various cold responses are not well understood. Our aim is to identify the epigenetic control mechanisms involved in cold acclimation and vernalization in *Brachypodium distachyon*. We applied several sequencing tools to seedlings subjected to cold for various amounts of time. Using data from these experiments we are characterizing the genome-wide changes in histone

modification, DNA methylation and RNA expression in order to develop testable hypotheses about the role of epigenetic regulation in cold responses.

P1124: Wheat, Barley, Oat, and related

Quantitative Trait Locus Analysis for Seed Metabolites Accumulation in *Brachypodium distachyon*

Yoshihiko Onda[1], Yuji Sawada[1], Yukiko Uehara-Yamaguchi[1], Masami Yokota Hirai[1], David F. Garvin[2], Kazuo Shinozaki[1] and Keiichi Mochida[1], (1)RIKEN CSRS, Yokohama, Japan, (2)USDA-ARS, Plant Science Research Unit, St. Paul, MN

Metabolite accumulation influences agriculturally and economically important traits of crops, such as grain quality and quantity. To investigate differences of metabolites accumulation in seeds among natural accessions in *Brachypodium distachyon*, we performed a widely-targeted metabolome analysis that detects known 517 compounds. Through the metabolome analysis of the natural accessions, we found that accumulation patterns of metabolites are significantly diverse among the accessions. Next, to identify quantitative trait loci (QTLs) associated with metabolites accumulation in seeds of *B. distachyon*, we assessed metabolites accumulation and identified at least 28 metabolites that are differentially accumulated between Bd21 and Bd3-1. We then used a series of 164 recombinant inbred lines (RILs) derived from a crossing between these accessions, and also used 551 single nucleotide polymorphism (SNP) markers for our metabolites QTL analysis. With the R/qtl software, we identified 6 metabolites that show high logarithm of odds (LOD) scores (LOD > 5). Specifically, we identified a QTL that is significantly associated with vitamin B_6 related metabolites, which is allocated on the *B. distachyon* chromosome 3. We further explored genes under the QTL peaks based on their functional annotations, and identified a candidate gene for this QTL. We also compared nucleotide sequences of the gene between parental accessions Bd21 and Bd3-1, and found a mutation of nonsynonymous substitution. These results suggest that this gene and its nonsynonymous change may be the cause for differential metabolite accumulation related to vitamin B_6 related metabolites between Bd21 and Bd3-1.

P1125: Wheat, Barley, Oat, and related

Genome Wide Selection for the Economic Value of Forage Yield in Tetraploid Perennial Ryegrass.

Stephen Byrne[1], Dan Milbourne[1] and Patrick Conaghan[2], (1)Teagasc, Crop Science Department, Carlow, Ireland, (2)Teagasc, Grassland Science Research Department, Carlow, Ireland

Genetic gain in forage species has been low in comparison to other crops. Reasons include a long selection cycle, an inability to select for "harvest index", an inability to exploit heterosis, and less investment. Indirect selection using genome-wide markers offers an opportunity to accelerate genetic gain in forage species by enabling multiple cycles of genomic selection (GS) to be performed in the same time it takes to perform a single cycle of conventional selection. We evaluated the accuracy of DNA based prediction for forage yield using a small training population that was free from population structure and had linkage disequilibrium that extended over multiple cMs. Even with the moderate prediction accuracies determined it should be possible to more than double the rate of genetic gain for forage yield.

P1126: Wheat, Barley, Oat, and related

Survey Sequencing of Flow-Sorted *Haynaldia villosa* Chromosome 6VS

Wentao Wan, Jin Xiao, Mengli Li, Haiyan Wang and Xiue Wang, Nanjing Agricultural University, Nanjing, China

Haynaldia villosa has been recognized as a useful germplasm for wheat breeding and improvement and the availability of genomic sequence would accelerate its research and application. In the present work, the short arm of *H. villosa* chromosome 6V in which powdery mildew resistant gene Pm21 have been mapped was flow-sorted by flow cytometry from a telocentric chromosome addition line of 6VS and sequenced using Illumina platform. We obtained a total of 47.7Gb raw sequencing reads and by de novo assembly 230.39Mb assembled sequence. Repetitive elements account for about 74.91% of the genome. 3,276 genes were annotated in the coding fraction of the genome which account for about 2.1%. The syntenic regions of 6VS genes were searched and identified on wheat group 6 chromosomes 6AS, 6BS, 6DS, rice chromosome 2, Brachypodium chromosome 3, and sorghum chromosome 4. Based on the size difference of intron for the synteny genes among 6VS genome and wheat group 6 chromosomes, we designed 222 IT markers, in which 120 markers had specific amplification on 6VS genome.The preliminary genomic sequence of 6VS provides genetic information for cloning genes on this chromosome and developing IT markers for molecular marker assisted breeding and physical map construction.

P1127: Wheat, Barley, Oat, and related

Editing of the *MeSWEET10a* Gene to Promote Tolerance to Cassava Bacterial Blight

Ihuoma Okwuonu[1], Kira Veley[2], Raj D. Chauhan[2], Joseph Onyeka[1], Chiedozie Egesi[1], Nigel J. Taylor[2] and Rebecca Bart[2], (1)National Root Crops Research Institute (NRCRI), Umudike, Umuahia, Nigeria, (2)Donald Danforth Plant Science Center, St. Louis, MO

The cassava *MeSWEET10a* gene has been associated with the virulence of *Xanthomonas axonopodis* pv. *manihotis* (Xam), the causal agent of cassava bacterial blight. Xam induces *MeSWEET10a* expression through the delivery of a type III effector protein, transcription activator like (TAL) 20, into the host cell that directly binds the *MeSWEET10a* promoter. We designed a strategy for developing tolerance to CBB by disrupting the TAL20 binding sequence, using the CRISPR-Cas9 gene editing technology. This strategy was evaluated in two cassava cultivars; TME 419 and 60444. Two guide RNAs (gRNAs) were designed to target the TAL20 binding site. After *Agrobacterium*-mediated delivery of the CRISPR-Cas9 reagents into cassava friable embryogenic callus (FEC), a total of 21 (cv. TME 419) and 7 (cv.60444) plant lines were recovered and analyzed for mutagenesis. Both gRNAs induced mutations independently at their target sites with deletion, insertion and substitution observed. Fifty-five percent (11/20) of lines recovered from TME 419 were shown to be heterozygous while thirty percent (6/20) were homozygous. For 60444, 67% (4/6) and 33% (2/4) showed heterozygous and homozygous mutations, respectively. Homozygous mutation within the TAL 20 binding site was not observed in events recovered from both cultivars and could be attributed to the reduced tolerance exhibited by these events. Future research will investigate the involvement of other factors in Xam virulence in cassava as well as the potential importance of *MeSWEET10a* gene in different developmental stages of cassava.

P1128: Other Plant Species

Phylogeny-Based Classification of Plant LTR Retrotransposons Fills the Gap between Superfamily and Family Levels and Provides Precise Annotation of Genomic Elements

Pavel Neumann[1], Petr Novak[2], Nina Hostakova[2] and Jiri Macas[1], (1)Biology Centre CAS, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic, (2)The Czech Academy of Sciences, Biology Centre, Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic

Plant LTR retrotransposons are classified into two superfamilies, Ty1/copia and Ty3/gypsy. Based on the overall DNA sequence similarity, both superfamilies are sub-classified into enormous number of families which are, due to high level of diversity, usually specific for a single species or a group of closely related species. Major disadvantage of the classification on the family level is the absence of any phylogenetic information. Relatively high similarity of protein domain sequences makes them ideal for inference of phylogenetic relationships, however, their thorough analysis across wide range of elements and species has not been done yet. Here, we analyzed a total of 75,516 polyprotein domain sequences identified in 13,863 LTR retrotransposon sequences extracted from 80 *Viridiplantae* species. Based on the sequence similarity and the order of protein domains in polyprotein sequences, 5410 and 8453 elements were classified as Ty1/copia and Ty3/gypsy, respectively. Phylogenetic analyzes of RT, RH and INT sequences allowed for finer classification of Ty1/copia and Ty3/gypsy retrotransposons into 16 and 14 lineages, respectively. In many cases, the classification was supported by lineage-specific features, including type and position of extra domains, presence, position and orientation of extra open reading frames, and the type of primer binding site. Whereas distinct lineages of Ty3/gypsy elements appeared to have relatively clear evolutionary relationships, Ty1/copia elements showed a complex evolutionary pattern involving events of recombination and horizontal transfers. The database of protein domains from the classified elements is implemented in the RepeatExplorer web server (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/) where it serves for automatic classification of transposable elements.

P1129: Other Plant Species

Biodiversity of Herbal Plants and Development of Species Authentication System using Low-Coverage Whole Genome Sequencing Data

Hyun-Seung Park, Jee Young Park and Tae-Jin Yang, Seoul National University, Seoul, South Korea

Plants are the sources of diverse natural compounds. Various natural products or medicines have been developed from herbal plants. However, most of the medicinal plants are non-model plants with a few basic genomic research and their complicated taxonomical classification system often causes mis-authentication with their counterfeit. The economically motivated adulteration of these plants based herbal products have been a serious threat for the consumer and the market. Therefore, we developed a new pipeline for analyzing low coverage Next-generation sequencing (NGS) data and applied this pipeline to herbal and resource plants in order to obtain fundamental genetic information and set up the basic authentication system. Using this pipeline, *de novo* assembly of low-coverage whole genome sequence (dnaLCW), complete chloroplast, mitochondrial genome and 45s ribosomal DNA could be produced simultaneously from the NGS data around 1x or even less coverage of the genome. Since 2015, more than three hundred of plant chloroplast and 45s rDNA belonging 69 genus and 33 family have been completed and several mitochondrial genomes were also assembled with this method. After comparative genome analysis with related species, their phylogenetic and evolutionary relationship could be revealed and molecular barcoding markers for authentication were also developed from the polymorphisms in these genomes. Besides, we developed an additional pipeline for mining polymorphic nuclear simple sequence repeat using same NGS dataset and applied this result for intraspecies diversity. This work was supported by the Bio & Medical Technology Development Program of the NRF, MSIP(Grant No. NRF-2015M3A9A5030733), Republic of Korea

P1130: Other Plant Species

Host-Induced Gene Silencing Improves Resistance Against Fusarium Oxysporum

Shulang Fei[1], Jonathan Peters[1], Elizabeth Aitken[2] and Neena Mitter[1], (1)Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Australia, (2)The University of Queensland, Brisbane, Australia

Fusarium oxysporum (Fo), a soil-borne fungal pathogen, is pathogenic to many crop species, causing root rot or wilt symptoms. Particularly banana, susceptible to *Fusarium oxysporum* f.sp. *cubense* (Foc), is facing huge challenge to survive in the overwhelming spread of the most aggressive Foc strain, tropical race 4 (TR4). Unfortunately, effective control methods to combat Fo remain limited. This project uses an RNA interference (RNAi) strategy, Host-Induced Gene Silencing (HIGS), to confer durable resistance against Fo. The rationale of HIGS is to introduce dsRNAs of particular Fo gene sequences into the plant so that the plant RNAi machinery is activated upon infection to inhibit the expression of targeted essential Fo genes, resulting in reduced virulence of the pathogen. To date four conserved regions from Fo gene coding sequences have been chosen as HIGS targets. Fo *in vitro* culture treated with dsRNAs corresponding to these regions showed significantly inhibited growth. Fo inoculation assay on transgenic Arabidopsis lines expressing these dsRNAs presented significant improvement with regards to survival rate and reduced wilting symptoms. The generation of the expected functional siRNAs was detected by small RNA northern blot. The *in planta* expression level of the targeted genes will be tested to confirm the resistance is conferred by the RNAi mechanism as hypothesized. A Red Fluorescence Protein (RFP)-expressing Fo strain will also be employed to investigate the Fo infection patterns in the transgenic lines. As the chosen target sequences are highly conserved among a series of Fo formae speciales which are the cause of a range of host crops, the HIGS system established in this study is expected to confer resistance in all of these hosts.

P1131: Other Plant Species

Safety Assessment of Transgenic High Energy Ryegrass to Improve Pasture Industries

Paula Giraldo, Department of Economic Development, Jobs, Transport and Resources, Melbourne, Australia

Perennial ryegrass is one of the most important pasture grasses in temperate zones worldwide. Genetic modification (GM) has been used to increase the metabolisable energy content of perennial ryegrass plants, through up-regulation of fructan biosynthesis in leaf blades. Commercialization of new transgenic cultivars requires data to support the deregulation of the transgenic event and this data considers the

impact of the novel germplasm on humans, animals and the environment. One aspect of this data is the development of tools to detect the transgenic event in relevant agricultural products.
A novel approach has been developed to evaluate high-energy transgenic ryegrass using emerging technologies, such as digital droplet PCR (ddPCR) and long read DNA sequencing. ddPCR enables absolute quantification of DNA copies to be determined, which allows detection of the transgene in agricultural commodities such as seed, herbage and silage at low levels, with accurate detection at 0.5%. Long read sequencing technology has also been applied to characterize the structure and flanking sequences of the transgenic insertion, through sequencing the transformed plant at moderate genome coverage. Further assessments have also been performed including toxicological and allergenicity studies indicating substantial equivalence to untransformed ryegrass. Nutritional trials using grazing animals are planned, which will validate the on farm production benefits. These novel technologies to assess and trace GM forages are being used to meet regulatory requirements to enable commercialization of this new, potentially large and transformative industry.

P1132: Other Plant Species

Pre-Breeding in Perennial Ryegrass (*Lolium perenne* L.) – a Nordic/Baltic Public-Private Partnership (PPP) Project

Odd Arne Rognli, Norwegian University of Life Sciences, Aas, Norway, Rene Aavola, Estonian Crop Research Institute, Jõgeva, Estonia, Torben Asp, Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, Gintaras Brazauskas, Institute of Agriculture, Lithuanian Research Centre for Agriculture and Forestry, Akademija, Lithuania, Aslaug Helgadottir, Agricultural University of Iceland, Reykjavík, Iceland, Mika Isolathi, Boreal Plant Breeding Ltd., Jokioinen, Finland, Anders S. Larsen, DLF Research Center, Store Heddinge, Denmark, Petter Marum, Graminor AS, Hamar, Norway, Christer Persson, Lantmännen SW Seed, Svalöv, Sweden and Sarmite Rancane, Latvia University of Agriculture, Jelgava, Latvia

Lolium perenne has superior feed quality and productivity, and is well adapted to the coastal climate of Western Europe. In Northern Europe, the climate change will probably improve the conditions for biomass production due to extended growing seasons (1-3 months) with milder and rainier autumns and winters. Thus, we expect perennial ryegrass to be grown further north under continental conditions. The main challenges is its susceptibility to snow moulds, and inadequate growth cessation in the autumn, with lack of sufficient cold hardening and winter survival as the result. Changing temperature-photoperiod regimes following the the new climates with higher temperatures are affecting plant phenology, winter survival and seasonal yield distribution. The genetic diversity available in the current Nordic/Baltic perennial ryegrass germplasm is restricted since the species is not native to the northern and continental regions. To address this challenge, a long-term Public-Private Partnership (PPP) on pre-breeding of perennial ryegrass started in 2012, involving academic partners and plant breeding companies from all Nordic and Baltic countries. The project aims at improving winter hardiness, persistence and other important traits for northern Europe by; i) investigating the current adaptation potential of commercial cultivars; ii) collecting and documenting plant material with large genetic variation from several parts of the world; iii) creating breeding populations with large variation and populations selected for extremes of different traits; and iv) developing tools for using genomic based selection in further breeding programmes. This presentation describes the experimental structure and design of this long-term pre-breeding project, and presents preliminary results.

P1133: Other Plant Species

Genes and Genetic Loci Related to Nitrogen Use Efficiency in Bermudagrass (*Cynodon dactylon*)

Raquel Schneider-Canny[1], Yun Kang[1], Patricio R. Munoz[2], Michael K. Udvardi[1] and Malay C. Saha[1], (1)Noble Research Institute, Ardmore, OK, (2)University of Florida, Gainesville, FL

Bermudagrass is the most important warm-season pasture in the southern USA with exceptional persistence and abiotic stress resistance. However, it requires high nitrogen (N) supply to reach its full biomass and quality potential. The objectives of this study were to determine the N use efficiency (NUE) of natural variants of bermudagrass, identify molecular markers associated with NUE related traits and validate underlying gene(s) and genetic mechanisms. An association mapping population consisting of 290 genotypes from the bermudagrass core collection and plant introductions from Germplasm Resources Information Network was established in the Ardmore, Oklahoma field with low inherent N. Phenotypic data on biomass yield and related traits, shoot N content and vegetative indexes were collected for two years. The population was also genotyped following the genotyping by sequencing protocol and SNPs were called based on ploidy groups. A total of 4,860 and 2,000 markers were identified for the triploid and tetraploid bermudagrass genotypes, respectively, with a cutoff at 20% missing values. In order to obtain additional markers for a more precise GWAS analysis, exome capture of the GWAS panel is in progress. Additionally, six bermudagrass genotypes with contrasting NUE were selected for transcript profiling under sufficient and deficient N growth conditions through RNA-seq. The primary differentially regulated genes identified include multiple leucine-rich repeat protein kinase family proteins, transcription factors, as well as genes involved in nitrogen uptake, protein/amino acid metabolism and transportation. The results obtained are being utilized to improve NUE through genetic and molecular approaches in bermudagrass.

P1134: Other Plant Species

Genome-Wide Association Analysis of Freezing Tolerance in Perennial Ryegrass

Andrius Aleliūnas[1], Cristiana Paina[2], Gražina Statkevičiūtė[1], Torben Asp[2] and **Gintaras Brazauskas**[1], (1)Institute of Agriculture, Lithuanian Research Centre for Agriculture and Forestry, Akademija, Lithuania, (2)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

Perennial ryegrass (*Lolium perenne* L.) cultivars have a moderate tolerance to freezing temperatures which limits their persistence at Northern latitudes due to poor overwintering. Genetic resources readily available from gene banks have vast collections of perennial ryegrass accessions and most likely include accessions of enhanced freezing tolerance (FT) and better winter hardiness. Here we report results from ongoing Private-Public Partnership on Pre-breeding in Perennial Ryegrass project where we have screened 150 gene bank accessions of perennial ryegrass for freezing tolerance under artificial freezing conditions. We demonstrate that ploidy level of perennial ryegrass determines the tolerance to freezing temperatures with diploid accessions being superior in FT to the tetraploid genotypes. The study has also revealed that natural genetic diversity for FT is already well exploited in some of the existing cultivars suggesting novel combinations of FT alleles are needed for further improvement. A subset of 122 diploid accessions was genotyped by sequencing resulting in 1.2 M Genome Wide Allele

Frequency Fingerprints. Genome-wide association analysis revealed a total of eight significant markers with a bias towards genic/expressed genomic regions. A number of candidate genes previously associated with abiotic stress response in plants were located on marker-tagged genomic scaffolds, while some of the tagged yet uncharacterized genes once validated might clarify the complex genetic mechanisms underlying FT in perennial ryegrass.
Acknowledgement: Funding provided by the Research Council of Lithuania (grant no. MIP-64/2015) and the the Nordic Council of Ministers.

P1135: Other Plant Species
Characterization of Imazamox Resistance in Jointed Goatgrass (*Aegilops cylindrica* Host)
Jeanette A. Rodriguez, Washington State University, Pullman, WA
Wheat growers have limited herbicide options to manage jointed goatgrass (*Aegilops cylindrica* Host), with many relying on imazamox or mesosulfuron in combination with Clearfield™ winter wheat. Both imazamox and mesosulfuron inhibit acetohydroxyacid synthase/acetolactate synthase (AHAS/ALS). In 2015, a suspected imazamox resistant biotype of jointed goatgrass was found in eastern Washington. Due to a shared D genome, jointed goatgrass (2n=CCDD=28) may outcross and hybridize with wheat (2n=AABBDD=42). Jointed goatgrass may acquire *ALS* resistance by hybridizing with Clearfield™ wheat, but no Ser_{653}(*At)*N mutation on the D genome of the resistant biotype of jointed goatgrass was identified. Sequencing efforts in the *ALS* gene indicate an Ala_{122}Thr substitution in the herbicide binding region of the *ALS* gene on the D-genome of jointed goatgrass. Increasing concentrations of imazamox were applied to both the resistant and a comparison susceptible jointed goatgrass biotype to test for resistance. The resulting dose-response data were fit using a 3-parameter log-logistic with GR_{50} (50% growth reduction) as a parameter. The suspected resistant biotype had a GR_{50} of 642 g ai ha^{-1} that is 29 times more resistant to imazamox than the known susceptible biotype with a GR_{50} of 22.21 g ai ha^{-1}. A field use rate of 12.3 times the recommended rate of 52.5 g ai ha^{-1} would be necessary to achieve 50% reduction in biomass. Thus, the Ala_{122}Thr substitution on the D genome of jointed goatgrass appears to confer a high level of resistance to imazamox. Future work will exclude alternate mechanisms and inheritance of resistance.

P1136: Other Plant Species
Genetic Structure of Native and Exotic Reed Canarygrass Populations
Andrzej Noyszewski[1], Neil O Anderson[2], Alan G. Smith[1], Diana Dalbotten[3], Emi Ito[4], Mike Dockry[5], Anne Timm[6] and Holly Pellerin[7], (1)University of Minnesota, Department of Horticultural Science, St. Paul, MN, (2)University of Minnesota, Department of Horticultural Science, Saint Paul, MN, (3)University of Minnesota, St. Anthony Falls Laboratory,, Minneapolis, MN, (4)University of Minnesota, Department of Earth Sciences, Minneapolis, MN, (5)USDA Forest Service, The Strategic Foresight Group, St. Paul, MN, (6)USDA Forest Service, Northern Research Station, Baltimore, MD, (7)University of Minnesota, St Anthony Falls Laboratory, Minneapolis, MN
Reed canarygrass *(Phalaris arundinacea)* is a wind-pollinated, wetland grass, cultivated in temperate regions around the globe as a forage and ornamental crop. It is also used for soil stabilization, bioremediation and bioenergy. In North America, it exists as both native and introduced populations from Eurasia that became invasive in MN. Native populations were identified from plants collected prior to 1940 in herbaria. The purpose of this study is to examine populations (along rivers and transects) in MN for their native, invasive vs. hybrid status to aid in land management. Reed canarygrass populations for genetic studies includes a total of 1,110 genotypes for analyses: 3 forage; 3 ornamental; 76 wild populations of 863 genotypes sampled every 30 km along 6 rivers (Mississippi, Minnesota, St. Croix, Red, Des Moines, Roseau) where at each site, 3 genotypes were collected at 10m intervals along each river edge as well as perpendicular transects; 166 genotypes from 6 transects in 3 wet meadows or cultivated fields collected along perpendicular, intersecting transects and historic 13 University of Minnesota (UM) herbarium samples that are native genotype. Genetic variation among and within reed canarygrass populations will be assessed by genotyping by sequencing (~1000 SNP markers) with support from a shallow genome sequence of a selected low alkaloid cultivar. It is critical to identify the extent of native vs. exotic reed canarygrass populations in MN using genetic testing for better management of exotic invasive populations and preservation of native populations in state and Tribal lands.

P1137: Other Plant Species
Sorghum Hybrids and their Parental Genetic Differences in Diversity
Xiaoping Lu, Inner Mongolia Agricultural University, hohhot, China
Objective. In order to further prove the diversity of genes in the leaves of hybrid and parental leaves of sorghum. Method. Four copies of the sorghum sterile lines and five types of Sudan grass for parents, according to the NC II design were formulated to 20 hybrid combinations. Measuring the cross combinations and field shape index and physiological and biochemical index of the parents.Using cDNA-AFLP differential display analysis of gene diversity difference. Result. The results of 29 sorghum hybrids and their parental cDNA-AFLP markers for the hybrids and their parental leaves were displayed:12 pairs of primer combinations amplified 336 sites, including specific sites 276, accounting for 82.14% of the total, an average of primer combinations amplified specific sites 23.The number of alleles (No) was 1.8701, the effective number of alleles (Ne) was 1.7365, the Nei's gene diversity index (H) was 0.3898, and the Shannon's (I) was 0.5569, which indicated that the gene expression of the hybrid and its parents were complex.The results of cluster analysis show that the change of Genetic similarity coefficient between 29 sorghum sudanense materials is between 0.508-0.876, and the Genetic similarity coefficient 0.54 is the standard, and the 29 materials can be divided into 4 categories.The parental materials were mainly in Iand the IV class, and the hybrids were mainly in the II and III. Conclusion. The reactive system of cDNA-AFLP differential display technique was constructed, and the genetic diversity of the leaves in the seedling stage was analyzed.
Key words: leaves of sorghum, cDNA-AFLP, heterosis, diversity of genes

P1138: Other Plant Species
Array-Based Comparative Study of Apomictic and Sexual Eragrostis Curvula Genotypes
Juan Pablo Selva[1,2], Ingrid Garbus[2,3], Cielo Pasten[2], José Carballo[2], Diego Zappacosta[4], Marco Pellino[5], Timothy Sharbel[5] and **Viviana Echenique**[4], (1)Dpto de Biologia Bioquímica y Farmacia, UNS, Bahia Blanca, Argentina, (2)CERZOS-CONICET,

Bahía Blanca, Argentina, (3)Departamento de Ciencias de la Salud - Universidad Nacional del Sur - Argentina, Bahía Blanca, Argentina, (4)Dpto. Agronomía and CERZOS/CONICET, Bahía Blanca, Argentina, (5)Global Institute for Food Security, Saskatoon, SK, Canada
Apomixis is an asexual reproductive process that results in seeds containing maternal clones since meiotic reduction and egg cell fertilization does not occur. Apomixis is common in polyploid grasses, and poorly represented in crop species of economic interest. A key goal in understanding the molecular basis of apomixis is the possible transference of this trait to species of agronomic relevance. Weeping lovegrass (Eragrostis curvula [Schrad.] Nees) is a perennial grass native to Southern Africa that reproduces by diplosporous apomixis. A custom 60-mer spots Agilent array was designed with 970 k probes based on a reference transcriptome constructed from inflorescence of sexual and apomictic genotypes. Eight independent hybridizations were performed with one-color Cy3 labelled samples (four apomictic and four sexual E. curvula genotypes). Data preprocessing and analysis was performed using the GeneSpring software v. 14.5. Platform quality was verified by Agilent spike-in controls. Normalization procedures consisted on percentile shift 75, background correction, median as baseline and reproductive mode as parameter. Individual probes from the array were considered to be differentially hybridized under the log2 transformed Fold Change data > 1 and p-values > 0.01 using unpaired t-test and multiple Benjamini-Hochberg correction. From these analyses 138 differential 60-mer sequences were obtained and four candidate genes strongly expressed in apomictic plants were identified. These genes could be have an important role in apomeiosis induction in weeping lovegrass and its relevance is currently being analysed through several complementary approaches, including in situ hybridization and obtention of Arabidopsis transformant.

P1139: Other Plant Species

A Genomic Approach to Study Apomixis using *Eragrostis curvula a*s a Model Species

José Carballo[1], Ingrid Garbus[1], Juan P. Selva[1], Bruno A. Santos[2], Emidio Albertini[3], Mario J Caccamo[2] and **Viviana Echenique**[1], (1)CERZOS CCT_CONICET Bahía Blanca, Bahía Blanca, Argentina, (2)NIAB, Cambridge, United Kingdom, (3)University of Perugia, Perugia, Italy
Apomixis is defined as asexual reproduction by seeds, avoiding meiotic reduction and fertilization, being generally present in polyploid plant species. *Eragrostis curvula* is a perennial grass native to Southern Africa. This species can be taken as a model for the discovery of genes that govern pseudogamous diplosporic apomixis since its polyploid cytotypes (4x to 8x) may undergo sexual reproduction, facultative apomixis, or obligate apomixis whereas diploids are always sexual.
Here we present the first draft of a diploid version of the *E. curvula* genome. The cultivar selected was Victoria (~1200 Mb) originated from *in vitro* culture of inflorescences of the apomictic cv. Tanganyika (2n=4x=40). Two libraries were prepared with fragment lengths of 20 kb and 10 kb to get longer reads and to increase the coverage, respectively. The sequencing through PACBIO technology resulted in 6.223.627 and 3.309.811 reads respectively with 90X coverage. The assembly was performed using the software Falcon. The N50 was 380.026 bp with 3.118 contigs representing 95% of the haplotype length. The software BUSCO was used to find single copy orthologous genes, being represented 97% of the BUSCO genes. Dovetail Hi-rise software revealed the architecture of the complete genome chromosome-by-chromosome.
The first draft of *E. curvula* genome showed high level of contiguity with a coverage of 95% of the diploid genome. The high proportion of annotated genes would allow the identification of those related to the reproductive mode. This draft represents the start point to obtain more complex tetraploid genomes, harboring the region/s involved in apomixis.

P1140: Other Plant Species

Molecular Characterization and Diversification Study in Wheatgrass Germplasms

Konstantin Chekhovskiy, Shyamal K Talukder, Michael Trammell, Dusty Pittman and Malay C. Saha, Noble Research Institute, Ardmore, OK
There is been a growing interest in improving native mono stands of Wheatgrass (*Triticeae trb.*) varieties in arid low nutrient marginal steppe lands. We have established two populations derived originally from unknown species of Mediterranean wheatgrass collections. Over the years of crossing within populations, final perennial hybrid cultivar “Enigma” was developed based on number of outperforming characteristics like drought tolerance, proliferation, persistence, biomass and forage quality. However, molecular characterization of the cultivar compare to other closely related known species and existing varieties is essential for enhanced improvement of the crop. A set of 48 SSRs previously reported in intermediate wheatgrass was used in this study. Phylogenetic analysis was done using free tree and tree view software’s. Enigma was found closer to other advanced germplasm, while diversity of the population pool was wide. The population was characterized for flowering and leaf characteristics. Molecular characterization will be continued and markers will be analyzed for finding important trait association. Molecular characterization of the variety will accelerate the breeding program for further improvement of wheatgrass.

P1141: Other Plant Species

Genomic Selection in Timothy (*Phleum pratense* L.)

Mallikarjuna Rao Kovi[1], Petter Marum[2], Muath Alsheikh[3], Helga Amdahl[2], Torben Asp[4] and Odd Arne Rognli[1], (1)Department of Plant Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Aas, Norway, (2)Graminor AS, Hamar, Norway, (3)Graminor, Ridabu, Norway, (4)Department of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark
Timothy (*Phleum pratense* L.) is a perennial, cross-pollinating hexaploid (2n = 6x = 42) species distributed naturally throughout Europe and parts of North Africa and Asia. It is one of the most important input factors in the Norwegian agriculture. New improved cultivars with high yield capacity and forage quality, which are well adapted to the future climate, are very important for an economically sustainable Norwegian milk and meat production. However, with traditional forage breeding methods it takes approximately 18-20 years to develop and release new cultivars on the market. Therefore, there is a great need for developing and implementing new breeding technologies in forage grass breeding. Genomic selection (GS) is a method that combines molecular markers with phenotypic and pedigree data for prediction of breeding values based on genome wide distributed markers. By selecting superior parents based on genomic information, the number of successful crosses increases and the breeding cycles shortens. In view of the recent developments in sequencing, molecular marker technologies and theoretical

foundations, it is feasible to start the development of GS based breeding schemes in timothy. The main objective of this project is to develop GS methods for breeding of the forage grass timothy by implementing the following strategies. 1) Developing single nucleotide polymorphism (SNP) marker allele frequencies (GWAFF) for 1000 full-sib families of timothy using genotyping-by-sequencing technologies. 2) Assessing the relative efficiency of GS for important timothy traits like yield and forage quality and trait stability across environments using different GS models. 3) Applying genome-wide association studies in conjunction with GS to unravel the genetic architecture of yield and forage quality. 4) Validating the GS predictions for yield and forage quality in unrelated 242 full-sib families.

P1142: Other Plant Species

Understanding Summer Dormancy and Marker Development in Tall Fescue using RNA-Seq

Shyamal K Talukder, Nick Krom, Konstantin Chekhovskiy, Jennifer Black, Yuhong Tang, Liang Sun and Malay C. Saha, Noble Research Institute, Ardmore, OK

Tall fescue is an important perennial forage grass in the USA. Mediterranean morphotype members of the grass adapts under harsh summer drought by exhibiting summer dormancy. Exploiting summer dormancy and devolve cultivar integrating the trait might be a potential solution for the southern great plains. Thus, understanding the mechanism in tall fescue and develop molecular marker has great importance. The study was conducted using one continental (R43-64) and two summer dormant (Agrafa 103-2 and GK45115-11) tall fescue. The genotypes were subjected under both summer dormant and optimum growing condition following two replications. Leaves, crowns and roots were collected for RNA-Seq at three different time point. De novo assembly was performed and 1,031,513 contigs were generated with average contigs size of 688.43 and N50 value of 988. Differential transcriptional response were found among the genotypes, samples and time points. SNP calling was performed and 386,000 SNPs were identified. By aligning the SNPs in Brachypodium genome 17,855 SNPs were found with annotation. A subset of 50 SNPs with annotation were validated using KASP.A total of 1100 SSRs with annotations were also validated using the same genotypes. Two SSRs were identified that discriminated Mediterranean and Continental genotypes. This study will lead to identify genes or transcripts involved in inducing and implementing summer dormancy in tall fescue which will eventually help understanding dormancy mechanism, as well as assisting the breeding efforts for integrating dormancy in tall fescue as a desired trait.

P1143: Other Plant Species

Genome-Wide Approach Based on Target-Enrichment NGS Sequencing Suggested Ancient Hybridizations in Magnoliaceae

Sangtae Kim, Sungshin University, Seoul, South Korea, Suhyeon Park, Dept. of biology, Sungshin Univ., Seoul, South Korea and Jongsun Park, InfoBoss Co., Ltd., Seoul, South Korea

Reconstruction of the phylogeny of Magnoliaceae, a member of early-diverging angiosperms, plays a key role in understanding the diversification of angiosperms. With the development of new molecular techniques, various phylogenetic studies on Magnoliaceae have been conducted such as studies based on chloroplast RFLP, and sequencing of single gene, multiple genes, and chloroplast whole genome. However, the evolution of Magnoliaceae remains unclear in many parts, especially the relationships among major sublineages in the family because overall base substitution rate in Magnoliaceae is very low, as compared to other angiosperm groups. In this study, we used a genome-wide approach on the phylogeny of Magnoliaceae with high throughput target-enrichment NGS sequencing. We identified 504 putative single/low copy gene regions (443 kbp) as targets for captured NGS sequencing. These regions are 1) expressed single-copy genes extracted from a preliminary genome assembly of *Magnolia kobus* (ver. 0.4; see talk 1401) using gene clustering (Tribe-Mcl) with its transcriptomes and 2) shared single/low copy genes among *Arabidopsis, Polulus, Oryza, Vitis*, and *M. kobus*. As a preliminary result, we determined chloroplast genomes (ca. 136 kbp excluding one IR) and 20 nuclear DNA regions (ca. 24 kbp) for 130 taxa representing all previously reported sublineages of Magnoliaceae. Chloroplast regions and nuclear regions produced robust phylogenetic trees. However, some conflict relationships were recognized between these trees. This suggests possible ancient hybridizations, especially between subgroups of MICHELIA and YULANIA clades. The results will provide a basis for the evolutionary diversification in the family as well as new classification system of Magnoliaceae.

P1144: Other Plant Species

Transcriptome Profiling of *Desmos chinesis*: Revealing the Molecular Basis of Dipartite Perianth Evolution in the Early Divergent Angiosperm Family Annonaceae

Amy Wing-Sze Leung[1], Sangtae Kim[2], Boon Leong Lim[1] and Richard M. K. Saunders[1], (1)The University of Hong Kong, Hong Kong, Hong Kong, (2)Sungshin University, Seoul, South Korea

Little is known of the genetic control of floral development in early-divergent angiosperms with differentiated sepals and petals. All Annonaceae species have flowers with a differentiated perianth, in contrast to the transitionally differentiated tepals observed in most other early-divergent angiosperms. In addition to a developmental study, we characterized the floral transcriptome and identified floral homeotic genes in *Desmos chinensis*, a representative of the Annonaceae, which shares the basic floral Bauplan of the family. In this study, transcriptomes of developing and mature *D. chinensis* floral organs and leaves were sequenced using the high-throughput Illumina platform. After filtering the coding regions (CDS), over 22,000 contigs were recovered from *de novo* assembly with N50 of 1407 bp. The identified CDS were annotated against five public databases. Based on sequence homology against NCBI nr database, CDS were annotated with 3,991 unique gene ontology (GO) terms and categorized into functional groups. GO enrichment analysis was performed to identify genes that are potentially significant for floral organ development. Transcripts that are commonly differentially expressed among developing and mature floral organs and leaves were identified using the digital gene expression data. We identified about 40 putative MADS-box gene homologs. Their respective expression levels were used to determine whether Annonaceae flowers are consistent with the MADS-box gene "fading borders" expression model. Selective MADS-box gene expression patterns were validated with qRT-PCR and *in situ* hybridization. While many questions remain unresolved at the transcript level, our data emphasizes the importance and potential of extending further studies into systems biology.

P1145: Other Plant Species

A Genome Survey of the Flowering Plant *Paris japonica* – the Largest Known Genome on Planet Earth

Depeng WANG[1], **Yang Meng**[1], Jingjing Li[1], Jiang HU[1], Jaume Pellicer[2], Ilia J Leitch[3] and Kai Wang[4], (1)Grandomics Biosciences, Beijing, China, (2)Royal Botanic Gardens Kew, London, United Kingdom, (3)Jodrell Laboratory, Richmond, United Kingdom, (4)Children's Hospital of Philadelphia, Philadelphia, PA
Paris japonica (Melanthiaceae) is a slow growing geophyte that is endemic to Japan. It is popularly known as the Japanese canopy plant because of its characteristic white umbrella-like flowers borne above a single whorl of stem leaves. From a genomic and biological perspective, it is particularly notable because it has the largest known genome so far recorded for any organism across all kingdoms of life (1C = 152.23 pg, i.e. ~150,000 Mb). The species is known to be octoploid, with 2n = 8x = 40 chromosomes, and is suspected of being an allopolyploid originating from four species. A detailed examination of its genome is likely to yield not only important insights into the compatibility of life with huge genomes, but also enhanced understanding of the dynamics and role of repetitive DNAs in shaping such giant genomes. Using DNA extracted from leaf material of Paris japonica from the Royal Botanic Gardens, Kew, we conducted genome skimming using Illumina short-read sequencing with 18X coverage (2,687 Gb data). The K-mer frequency analysis (i) revealed that the estimated genome size was ~152,145.645,170 bp, which is consistent with previous estimations based on flow cytometry, and (ii) generated a unimodal curve indicating a low rate of heterozygosity. We also analyzed the types and abundance of the most common repeat elements from the sequence data, and performed a preliminary genome assembly to infer possible functional elements. Based on our findings so far, we discuss and hypothesize why and how some plants can tolerate genome obesity, via multiple mechanisms including alloploidy, autopolyploidy, and expansion of specific classes of repeat sequences (including transposable elements), that may help to maintain chromosomal integrity during DNA replication.

P1146: Other Plant Species

Unravelling Chemical Diversity in Cannabis Using an Extreme Chemical Phenotype Genome-Wide Association Study

Matthew Welling[1,2], Lei Liu[2] and Graham J. King[1], (1)Southern Cross Plant Science, Southern Cross University, Lismore, Australia, (2)Southern Cross University, Lismore, Australia
Cannabis is a domesticated genus that produces a unique class of chemically diverse and therapeutically valuable secondary metabolites referred to as cannabinoids. Modulation of the human endocannabinoid system is influenced by the chemical structure of cannabinoids, and these structural-activity relationships are reliant on alkyl side chain length. It is anticipated that the characterisation of alkyl side chain variation at the genetic level will accelerate the selection of novel breeding lines for pharmaceutical exploitation. However, the genomic constituents responsible for alkyl side chain length are under-characterised, and the ability to predict alkyl cannabinoid composition at early developmental stages *in planta* remains limited. High-throughput DNA sequencing approaches have the potential to improve the accuracy of determining chemical diversity, although the suitability of these methodologies should be examined rigorously in non-model outcrossing heterozygous species such as *Cannabis*. In addition, these approaches should be measured by their ability to elucidate the complex genomic architecture associated with secondary plant metabolism. To enrich for functional genetic variants, we propose the use of a contemporary methodology which involves whole-genome sequencing of pooled extreme chemical phenotypes from a globally representative *Cannabis* diversity panel (sourced from the Ecofibre Industries Operations Pty Ltd genetic resource collection). We expect this approach to have more general application for secondary metabolite marker development, including for other medicinally relevant orphan plant species. For *Cannabis* we expect contributions to the genetic metabolic enhancement of alkyl cannabinoid composition for therapeutic end-use.

P1147: Other Plant Species

Finnish Hops Who Are You?

Lidija Bitz, Merja Hartikainen, Juha-Matti Pihlava, Pirjo Tanhuanpää, **Sirja Viitala** and Teija Tenhola-Roininen, Natural Resources Institute Finland, Jokioinen, Finland
Hop (*Humulus lupulus* L.) is climbing plant belonging to the *Cannabaceae* family. Hop cones are containing alpha acids and essential oils that give bitterness and unique aroma to the beer thus making hops widely used in brewing industry and commercially important. Hops are growing in Finland being native and have been cultivated over a century ago. With times commercial hop yards vanished due to the increased import of hops. In the last century there have been several trials to revive hop production but did not result in wider scale cultivation. Nowadays a popularity of local hops is growing rapidly together with only rising demand for crafts beer.
This clear need for locally produced hops inspired us to start a study of the Finnish hops. For the purpose characterization of hops growing all over Finland, we collected samples of over 800 putatively different accessions from old gardens and collections. We analyzed the contents of alpha- and beta-acids and prenylated flavonoids by liquid chromatography (HPLC-DAD), terpenoids by gas chromatography (GC-MS) and for accession with sufficient number of cones also the amount of volatile oils by steam distillation. Genetic analyses were performed with 25 microsatellite markers including also reference landraces and bred cultivars from Europe and rest of the world. We started analysis with a subset of collected hop accessions (200) and the results showed that Finnish hops clustered to a separate group distinguishable from the reference samples.

P1148: Other Plant Species

A Segregating Population of *Mentha longifolia* Enables Molecular Marker Development for Mint Oil Type, Male Fertility and Verticillium Wilt Resistance

Iovanna Pandelova[1], Saddie Vela[1], Amber Parrish[2], Iris Lange[2], Bernd Markus Lange[2] and Kelly Vining[1], (1)Oregon State University, Corvallis, OR, (2)Washington State University, Pullman, WA
Oil yield and quality are the most important traits for the marketability of mints. Black Mitcham, the most-widely grown peppermint cultivar, is susceptible to Verticilium wilt, caused by fungal pathogen *Verticillium dahliae.* Verticillium wilt-resistant mints exist in nature and, in the case of Native spearmint, in cultivation. Developing new peppermint hybrids which contain alleles for peppermint-type oil production and Verticillium wilt resistance would provide for the long-term sustainability of this crop. However, conventional breeding for Verticillium wilt resistance is currently impossible due to the sterility of peppermint. With the advent of high-throughput genome sequencing, we can now determine the sequences of genes correlated with desirable oil characteristics, wilt resistance and pollen fertility traits in mint. To address these, we are taking advantage of the South African F2 (SAF2) population of segregating lines derived from a cross of wild diploids, *M. longifolia*

USDA accessions CMEN585 and CMEN584. Our oil analyses indicated that the SAF2 population represents four major oil chemotypes (chemically distinct variants): 1) resembling parent CMEN585; 2) resembling parent CMEN584; 3) intermediate between both parents; and 4) different from either parent. All lines have also been tested for Verticillium wilt resistance and pollen fertility. Using genotyping-by-sequencing (GBS) data we found Single Nucleotide Polymorphism (SNP) markers associated with wilt resistance and pollen fertility. SSR markers for oil genes were developed using the *Mentha* reference genome sequence derived from CMEN585. Markers were screened in the SAF2 population utilizing High Resolution Melting analysis (HRM) and PCR.

P1149: Other Plant Species

Phylogenomics of the *Dioscorea* Yams, a Major Pantropical Crop

Marybel Soto Gomez[1], Juan Viruel[2], Paul Wilkin[2] and Sean Graham[1], (1)University of British Columbia, Vancouver, BC, Canada, (2)Royal Botanic Gardens, Kew, London, United Kingdom

The *Dioscorea* yams are the third most important tropical tuber crop globally. Multiple species in the genus have been brought into cultivation for their starch-rich annual tubers, providing a major source of nutrition and income to millions of subsistence farmers. Yams with perennial tubers and rhizomes have also been exploited for pharmaceutically useful compounds, like diosgenin. Although progress has been made in understanding *Dioscorea* relationships, multiple uncertainties persist, including the precise identity of crop wild relatives (CWRs) of major yam cultigens. A phylogenetic framework for the genus would be useful for identifying CWRs that may represent genetically similar pools for germplasm selection, and for future comparative genomic studies. We are addressing this knowledge gap by using targeted enrichment of nuclear genes from wild-collected species represented largely by herbarium material. We are using baits to target 303 nuclear genes (260 single- to low-copy genes, and 43 genes of functional interest) from ~400 species of *Dioscorea*, spanning the taxonomic diversity of the genus and including all major yam crop species and putative CWRs. We present our preliminary phylogenomic results here. We are also comparing the leaf and underground organ transcriptomes of cultivated and wild species to investigate the genes underlying traits of interest, such as starch and metabolite accumulation in the diverse underground organs in the genus, and to study the selective forces that may drive these differences. This will lay the groundwork for characterizing the genetic basis of differences in the genus, including those between yam crop species and their CWRs.

P1150: Other Plant Species

Construction of an Ultradense Genetic Map in Hexaploid Sweetpotato

Marcelo Mollinari[1], Bode Olukolu[1], Guilherme Da Silva Pereira[1], Dorcus C. Gemenet[2], Awais Khan[2], Mercy N. Kitavi[3], Marc Ghislain[3], Craig Yencho[1], Zhao-Bang Zeng[1] and David Maria[2], (1)North Carolina State University, Raleigh, NC, (2)International Potato Center (CIP), Lima, Peru, (3)International Potato Center (CIP), Nairobi, Kenya

The hexaploid sweetpotato (*Ipomoea batatas* (L.) Lam., 2n = 6x = 90) is an important staple food crop worldwide. Due to its complex genome, the construction of genetic maps in sweetpotato and other high polyploid organisms has been restricted to the use of simplex markers. We have constructed an ultradense sweetpotato integrated genetic map using simplex and multiplex markers in a biparental population derived from the cultivars “Beauregard” and “Tanzania”. The mapping population consisted of 315 individuals genotyped by a modified GBS protocol optimized for polyploids (GBSpoly) using the genome of a related diploid sweetpotato (*I. trifida*) as reference. A total of 90,418 SNPs were called from the raw data using SuperMASSA software. After filtering, this number was reduced to 20992 high quality SNPs (50% simplex (0 x 1), 10% double simplex (1 x 1) and 40% several multiplex configurations). The genetic map was constructed using the software polymap, which uses pairwise marker analysis in combination with hidden Markov models to infer the linkage phase and the recombination fraction. The resulting genetic linkage map consisted of 16,107 SNPs distributed in 15 homology groups with a total length of 1303.1 cM. We observed a marked colinearity between the *I. batatas* map and the *I. trifida* reference genome with noticeable rearrangements in homology groups 2, 3 and 7. Using the final map we obtained the conditional probability distribution for all possible 400 genotypic classes at every 0.5 cM which were used in QTL analysis in further studies.

P1151: Other Plant Species

Designing SNP-Array for Guinea Yams (*Dioscorea* spp.) for Routine Use in Breeding Program

Angelot Paterne AGRE[1], Ranjana Bhattacharjee[1], Andreas Gisel[1], Michael Abberton[1], Asrat A. Amele[2] and PAG author's group, (1)International Institute of Tropical Agriculture, Ibadan, Nigeria, (2)IITA, Abuja, Nigeria

Guinea yam (*Dioscorea* spp.) is a food security crop in the ‘yam-belt’of sub-Saharan Africa. Very recently, the reference genome sequence of guinea yam *(D. rotundata*) has been completed. Additionally, advanced molecular approaches including genotyping-by-sequencing has been completed in 941 accessions including cultivated guinea yam and wild relatives. Two different bioinformatics pipelines, Genome Analysis Tools Kit (GATK) and TASSEL ver 5 was used for SNP identification. Filtering of raw SNP data from the GATK pipeline based on minimum allele frequency (MAF), missing value percentage, and read depth resulted in 4,117 SNPs. To obtain a reasonable number of SNPs for routine use in breeding program of guinea yams, further analysis was carried out based on pruning linkage disequilibrium (LD), Fst, polymorphism information content (PIC), and MAF to obtain 192 filtered SNPs distributed across all linkage groups. These SNPs were validated across 285 diverse clones to test their polymorphic nature.

P1152: Other Plant Species

Towards Comprehensive Control of *Yam Mosaic Virus* in West Africa

Ranjana Bhattacharjee[1], P. Lava Paterne[1], Chukwuemeka K. Nkere[2], Beatrice Aighewi[3], Norbert Godonou Maroya[4], Joseph Onyeka[5], Asrat A. Amele[6], Michael Abberton[1], Susan E. Seal[7] and Robert Asiedu[1], (1)International Institute of Tropical Agriculture, Ibadan, Nigeria, (2)National Root Crop Research Institute, Nigeria, Indianapolis, IN, Nigeria, (3)IITA, Ibadan, Nigeria, (4)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (5)National Root Crops Research Institute (NRCRI) Umudike, Umuahia, Nigeria, (6)IITA, Abuja, Nigeria, (7)NRI, University of Greenwich, Chatham Maritime, United Kingdom

Yam mosaic virus (YMV, genus *Potyvirus*) is the most important viral pathogen infecting yams (*Dioscorea* spp.) in West Africa. Annual incidence of YMV in the two most widely cultivated species, *D. rotundata* and *D. cayenensis* exceed 70%, resulting in about 40% production loss per annum in West Africa. Perpetual use of farmer-saved seed has been recognized as the major contributing factor for high YMV incidence in the farmer fields. Systematic efforts since past 5 years have contributed to mapping of YMV spread and characterization of over 120 YMV isolates that revealed circulation of diverse strains, consequence of which on symptom types and host resistance response are yet to be understood. This knowledge however helped in development of versatile diagnostic tools for YMV indexing, and establishment of virus-free planting materials of popular cultivars necessary to invigorate 'clean' seed systems. Efforts to understand the rate of virus infection offered clues to YMV epidemiology and this knowledge has been employed for on-farm management to reduce YMV incidence and severity. More recent efforts are oriented towards characterization of yam germplasm for identification of YMV resistance in farmer adopted cultivars and also to select promising parents for YMV resistance breeding. Integration of multiple approaches is expected to result in the comprehensive management of YMV which is critical not only to recover yield lost due to virus infection, but also to double productivity to meet the growing demand for one of the main staple food crops in West Africa. This presentation seeks to provide diversity and distribution of YMV in various yam cultivars and geographies and implication of this knowledge on decisions on YMV control strategies.

P1153: Other Plant Species

Multiple QTL Mapping in Hexaploid Sweetpotato for Yield and Yield Components

Guilherme Da Silva Pereira[1], Dorcus C. Gemenet[2], Marcelo Mollinari[1], Bode Olukolu[1], Federico Diaz[2], Veronica Mosquera[2], Wolfgang Gruneberg[2], Awais Khan[2], Craig Yencho[1] and Zhao-Bang Zeng[1], (1)North Carolina State University, Raleigh, NC, (2)International Potato Center (CIP), Lima, Peru

Despite of the social and economic importance of several autopolyploid species such as sweetpotato, *Ipomoea batatas* (L.) Lam. (2*n*=6*x*=90), detection and characterization of quantitative trait loci (QTL) are still limited. Due to the genome complexity of autopolyploids, current single-QTL fixed-effect model fails to incorporate genetic terms other than the additive ones and are generally hard to interpret. Here, we used a random-effect model to map multiple QTL in a population with 315 full-sibs derived from a cross between the sweetpotato cultivars 'Beauregard' and 'Tanzania'. Phenotypic data were collected for 14 traits (including vine weight, and commercial, noncommercial and total root number and weight) in six environments in Peru, and mixed models were used to obtain joint adjusted means for each trait. An integrated linkage map consisting of 16,107 markers distributed along 15 linkage groups spanning 1,303.1 cM was used to obtain the genotype conditional probabilities at every 0.5 cM. Multiple interval mapping was performed using the R package polyqtl and detected a total of 53 QTL, with the number of QTL per trait ranging from 6 to 1. Some regions, such as those on chromosomes 1 and 15, were consistently detected among root number and weight related traits. In addition, some QTL were found to underlie commercial and noncommercial root traits distinctly. Best linear unbiased predictions allow us to characterize additive and higher order allele interaction effects as well as to compute the breeding value of the full-sibs for selection.

P1154: Other Plant Species

Comparative RNA-Sequencing Analysis of Storage Roots in Sweetpotato

Sathish Kumar Ponniah[1], Thiruvarangan Ramaraj[2], Venu (Kal) Kalavacharla[3] and Muthusamy Manoharan[1], (1)University of Arkansas at Pine Bluff, Pine Bluff, AR, (2)National Center for Genome Resources (NCGR), Santa Fe, NM, (3)Delaware State University, Dover, DE

Sweetpotato storage roots function in carbohydrate storage and vegetative propagation and form from adventitious roots. Adventitious roots develop from nodal primordia and cut ends or wounds of the stem (slips) after transplanting. These adventitious roots can then form storage roots by a process that involves thickening of the vascular tissue, followed by the accumulation of starch and proteins. The conversion of adventitious roots to storage roots involves the formation of new cambial cells with the development of secondary cambium and thin-walled parenchyma cells. However, the molecular mechanism of storage root formation is not clear. RNA-sequencing was carried out between roots of the cultivated storage root forming species (*Ipomoea batatas*) and its non-storage root forming wild ancestors (*I. triloba*, *I. cordatotriloba*, and *I. tabascana*) to understand storage root forming genes. A total of 66,418 genes were identified in the species. While 65.6%, 66.7%, and 68.5% unique genes were found in *I. triloba*, *I. cordatotriloba*, and *I. tabascana* respectively, and 7.2% unique genes were identified in the cultivated storage root forming species. The unique genes identified in the storage root forming sweetpotato species (*I. batatas*) include calcium-binding proteins (CML40), Glutathione S-transferase, Ferrochelatase, Cysteine protease, and Chorismate mutase. We are also examining gene expression through real-time PCR between the cultivated storage root forming species, and the non-storage root forming ancestors.

P1155: Other Plant Species

Genetic Diversity within CIAT's Cassava Germplasm Collection

Bruno A. Santos[1], Mohamed Abdelhalim[1], Pradeep Ruperao[1], David Marshall[2], Peter Wenzl[3] and **Sarah C. Dyer**[1], (1)NIAB, Cambridge, United Kingdom, (2)The James Hutton Institute, Dundee, United Kingdom, (3)CIAT, Cali, Colombia

CIAT's Genetic Resources Program houses >6,000 accessions of cassava (*Manihot esculenta*) and its wild relatives. Cassava is a globally important food crop and staple for >500 million people in Africa, Asia and South America, and is also used for feed, starch and biofuel. Cassava is high in carbohydrate, though relatively low in other nutrients; it grows well in poor soils and with unpredictable rainfall, making it a useful famine reserve crop whose cultivation is predicted to increase linked to changing climates. However, cassava suffers from susceptibility to a number of pests and pathogens, which can severely impact yield. The natural diversity contained within the genebank's collection may provide novel sources of biotic stress resistance and useful alleles linked to traits of interest.

Relatively little is known about the accessions within the genebank, aside from their collection information, with the majority of accessions coming from Colombia and Brazil. The cassava genome is diploid and ~770Mb in size. Cassava suffers from in-breeding depression and is clonally propagated, as such the samples are expected to be highly heterozygous. We have genotyped >4,000 accessions from the collection using DArT-Seq, generating >30,000 high confidence allele calls mapped to the reference genome. Using this genotyping data we have also

selected 25 genetically diverse individuals to be whole-genome sequenced to allow us to explore their genomic diversity more fully. All data will be available through a Germinate 3 instance currently under development within CIAT's genetic resources web portal.

P1156: Other Plant Species

Application of the Axiom 3K SNP Genotyping Array in Cassava Breeding and Genetics

Marco Pessoa-Filho[1], Pedro I.T. Silva[2], Lucileide V. Resende[3], Eduardo A. Vieira[1], Fabio G. Faleiro[1], Dario Grattapaglia[3,4] and Orzenil B. da Silva Junior[3], (1)Embrapa Cerrados, Brasilia, Brazil, (2)Department of Cell Biology - University of Brasília, Brasília, Brazil, (3)Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF, Brazil, (4)Graduate Program in Genomic Sciences - Catholic University of Brasília, Brasília, Brazil

Use of high-throughput SNP genotyping platforms improves technical quality and traceability control in breeding programs and germplasm management. We had previously reported the development of a 3,417 SNP array for *Manihot esculenta*, as part of a five-species 50K SNP Axiom genotyping array for cashew, coffee, araucaria, eucalypts and cassava (Grattapaglia et al. 2017). SNPs were selected from variants identified in resequenced genomes of 52 worldwide cassava accessions publicly available from Phytozome. A subset of 160,000 SNPs was initially screened from 26 million variants for optimal array design, targeting 16 South American wild and cultivated cassava and 4 *M. glaziovii* accessions. A list of 3,417 high-quality SNPs distributed throughout 18 chromosomes and scaffolds of the v6.1 AM560-2 genome assembly was sent to final design. In this work, this SNP array was used for pedigree reconstruction of open pollinated progeny and to characterize 447 accessions of the Embrapa Cerrados cassava germplasm collection. Maximum-likelihood parentage analysis determined the correct parents for 91 progeny individuals. For only 35 of them these matched the expected parents from pedigree records, clearly showing the advantage of using SNP data to guarantee the correct pedigree and thus the estimates of heritability and genetic gain thereof. Multi-locus identity analysis revealed a large number of duplicated accessions, such that only 204 of the 447 accessions in the collection were actually unique. Discriminant analysis of principal components and fastStructure indicated the collection was composed of six clusters. A naïve genome-wide estimate of linkage disequilibrium (LD) r^2 of 0.019 was significantly reduced when corrected for population structure and relatedness going to r^2_{VS} of 0.005, decaying to < 0.2 within *c.* 75 kb. These results offered a first simple glimpse at the huge impact that the integration of SNP data can cause in cassava breeding and genebank management. Given the very high data quality and plummeting costs of array genotyping especially in a multi-species format, we anticipate its routine adoption to accelerate breeding via genomic selection.

P1157: Other Plant Species

Status of Cassava Viruses in Northern-Nigeria

Ibrahim Umar Mohammed, Kebbi State University of Science and Technology, Aliero (KSUSTA), Nigeria, Aliero, Nigeria
I.U. Mohammed[1], A. Muhammad[1], M. N. Maruthi[2] and WAVE team members
[1]Kebbi State University of Science and Technology, Aliero (KSUSTA)
[2]Natural Resources Institutes University of Greenwich, UK

Capacity of farmers, plant breeders and virologists was developed in northern Nigeria. By setting-up of laboratory in KSUSTA, with capability to provides surveillance, diagnosis and characterizations of plant viruses in the region. Farmers, extension workers and quarantine officials were trained on cassava mosaic disease CMD and cassava brown streak disease (CBSD) symptoms identification. The status of CMD and population of whitefly was established by two surveys conducted in September 2015 and November 2016 in 13 States of the north-west and north-eastern Nigeria. CMD occurred in all the States surveyed, although its incidence was higher in Taraba State (86%) than in Kaduna (21%). Population of whitefly was higher in both regions averaging 10.3 adults per shoot in north-west and 6.5 in north-east. Most infections were attributed to the use of infected cuttings, 88 and 91% in north-west and north-east, respectively. Average disease severity was generally low in north-west (2.6) than in north-east (3.7). The highest disease severity was recorded in Taraba State (4.0) and the lowest in Kaduna (2.0). PCR work on leaf samples indicates the presence of African cassava mosaic virus (ACMV) and east African cassava mosaic virus (EACMV) as single and in mixed infections. The lack of awareness of the negative effects of these viruses signalled a great danger to the national food security, thus the need for capacity building in the region for the effective control of CMD in Nigeria.

Corresponding author: iumohammed74@gmail.com

Keywords: northern-Nigeria, Virus species, CMD, disease incidence, disease severity

P1158: Other Plant Species

Comparison of within and between Cluster Crosses: The Case of Cassava Brown Streak Disease Resistance

Emmanuel Frank Mrema, Mikocheni Agricultural Research Institute, Dar es Salaam, Tanzania, United Republic of
Emmanuel Mrema[1]*, Betty Owor[1], Robert Kawuki[2], Alfred Ozimat [2], Yona Baguma [2], Richard Edema[1], Paul Gibson[1]
[1]Collage of Agricultural and environmental Sciences, Makerere University, P.o. box 7062 Kampala, Uganda, [2]National Crop Resources Research Institute, P.o. box 7084 Kampala, Uganda

Abstract

Cassava brown streak disease (CBSD) is a major threat to cassava production in eastern, central and southern Africa countries, were it causes yield losses of up 100%. Different genotypes have been reported to express different resistance mechanism to CBSD, creating the need for generating varieties that combine these different resistance mechanisms. Thus, with help of CBSD field data and genotype data, we selected 52 parental clones for hybridization based on genomic estimated breeding value. These parental clones were grouped into four genetic clusters (based on SNP data). Thus crosses were made between and within clusters. The generated 378F_1 seedlings were established in the field for CBSD evaluations. The trial was set in one hot spot environment in 7 by 10 alpha lattice design where families were replicated twice. Result indicated that different cluster combination had different levels of CBSD resistance; in some case the progeny outperformed the parental clones. ANOVA showed significant difference ($P \leq 0.001$) for the 10 combinations from within and between cluster crosses. Further, t-test revealed significant heterosis at $P \leq 0.001$ from within and between cluster crosses. Further, highest negative heterosis was revealed associated with low CBSD severity mean score from within cluster crosses.

P1159: Other Plant Species

Breeding Schemes to Increase Purging of Deleterious Alleles in Cassava

Jean-Luc Jannink, USDA-ARS / Cornell University, Ithaca, NY and Mohamed Ibrahim, Cornell University, Ithaca, NY

Recent research has shown that elite cassava clones may harbor many recessive deleterious alleles masked in the heterozygous state. This finding is consistent with severe inbreeding depression in cassava and high levels of non-additive variation for root yield. We simulated cassava breeding schemes that involved a generation of selfing to expose deleterious alleles and increase the rate at which they might be purged. We will report on simulation results of rates of improvement that can be obtained from traditional cassava breeding schemes compared to schemes that incorporate selfing. Additionally, we report on the effect such purging schemes might have on the severity of inbreeding depression and the level of non-additive variation in cassava.

P1160: Other Plant Species

High Throughput Quantitative Phenotyping of Resistance to Whiteflies (*Aleurotrachelus socialis* Bondar) on Cassava Using Image Analysis

Adriana Bohorquez, International Center for Tropical Agriculture (CIAT), Cali, Colombia

The whitefly*, Aleurotrachelus socialis*, is one of the pest that cause bigger harm in agroecosystems, by feeding on the phloem of the leaves and transmit bacterial and fungus diseases, especially on cassava (*Manihot esculenta* Crantz). Therefore, it is important to develop alternatives phenotyping tools that streamline the process of finding genetically resistant varieties to this pest. We developed an automated phenotyping tool as a plugin to ImageJ (ImageJ, National Health Institute, Bethesda, USA), which efficiently quantifies the number of nymphs in third and fourth instar on cassava plants, in order to access to plant resistance level. Our data analysis showed a high correlation and a temporal reduction in more than 80% of evaluation time per plant, in comparison with conventional visual of counting method. Additionally, were added measures as: leaf's area and nymphs' area (area in pixels), and the percentage of affected leaf. Aiming to evaluate the evolution and growth of the plant and in that way provide a more robustness result for the study of resistance trait in plants. Automated nymphs quantifier allows to evaluate a large number of images for a plant populations in a short time, doing it accurate, more efficient and especially avoiding the variability and potential observer bias normally associated with manual counting. We will show the results of the plugin analysis on 11 important cassava genotypes for the study of whitefly resistance in this crop.

P1161: Other Plant Species

Whitefly Resistance QTL Mapping in Cassava: The Case of an Ecuadorian Resistance Source against *Aleurotrachelus socialis*

Adriana Bohorquez, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Whiteflies (WFs) are a major threat for cassava production, due the damage caused directly their phloem feeding behavior and indirectly by serving as vectors of viral diseases. Aleurotrachelus socialis is the most important WF species in Latin-America. Several M. esculenta cultivars originating from Ecuador and Peru (line ECU72) display significant resistance to WF in the field. In order to define the genetic regions involved in the resistance against A. socialis we used QTL mapping on two segregating populations derived from parents of contrasting resistance phenotype, CM8996 (ECU72 x COL2246) and GM8586 (ECU72 x NGA11). Choice experiments were carried out for assessing resistance in greenhouse conditions and nymphs were counted using computer vision software. Data was summarized though a generalized linear mixed model with negative binomial error and resistance was measured though the resulting best linear unbiased predictors (BLUPs). RADseq (CM8996) and whole genome sequencing (GM8586) methods were used for genotyping and the resulting markers were the base for building high density linkage maps. Interval mapping analysis resulted in the detection of 6 resistance QTLs in the 2 populations. A single QTL (LOD 3.45) in Chromosome 10 from CM8996 explains 18% of the BLUP variance observed in the different family assays. This percentage, corresponds most of the additive variance estimated by interexperiment reproducibility (maximum R2 = 0.27). For GM8586, in contrast, the greatest effect QTL (Chromosome 18, LOD 3.1) explains 8% of the BLUP variance, and interexperiment reproducibility is very low (R2 = 0.02).

P1162: Other Plant Species

Genome-Wide Association Studies of Branching and Flowering Traits in Cassava (*Manihot esculenta* Crantz)

Elohor Mercy Diebiru-Ojo, International Institute for Tropical Agriculture, Ibadan, Nigeria

Abstract

Flowering in cassava (Manihot esculenta Crantz) is complex and little is known about genetic diversity and inheritance of flowering traits and associated plant architecture patterns. Genetic studies of cassava flowering were conducted in Nigeria using the IITA genetic gain population of 750 genotypes in two locations for three years. Broad sense heritabilities for flowering and branching traits were estimated. The traits were further subjected to the genome-wide association studies (GWAS) to identify marker-phenotype associations. The set of 750 cassava genotypes were phenotyped and in parallel GBS-genotyped using the ApeKI restriction enzymes. SNPs were called and filtered generating 70414 informative SNP markers. The broad sense heritabilities estimated ranged from 33% - 75% for seven traits including days to anthesis, number of female flower, number of male flowers, number of fruitset, plant height, number of branch levels and first branch height. A total of 345 marker-trait associations (MTAs) were identified using a generalised linear model (GLM) which yielded 327 MTAs and the mixed linear model (MLM), 18 MTAs. The results provide a list of genomic loci which should be functionally validated. These data serve as a baseline for functional genomic studies of cassava and efforts to control cassava flowering to enhance efficient production of crosses and recombination.

Keywords: Flowering, Genomic, Broad sense heritability, MTAs, GWAS.

P1163: Other Plant Species

Transient Expression and Purification of Cassava Brown Streak Virus Coat Proteins in *Nicotiana benthamiana* for Development of Diagnostic Tests

Angela O. Eni[1], Keith Saunders[2] and George P. Lomonossoff[2], (1)Department of Biological Sciences, Covenant University, Ota, Nigeria, (2)Department of Biological Chemistry, John Innes Centre, Norwich, United Kingdom

Cassava is the preferred staple food for over 800 million Africans and over 32% of global cassava production takes place in West Africa. Sustainable and profitable cassava production is threatened by devastating Cassava brown streak disease (CBSD) which is currently widespread in East and Central Africa, but not present in West Africa. Although specific primers for the detection of *Cassava brown streak virus* (CBSV) exist, lack and/or inadequate infrastructure complicates the use of available molecular tests for routine CBSD monitoring in West Africa. To generate antigenic proteins useful for CBSV antibody production and non-infectious positive control useful in West Africa, we investigated *Agrobacterium*-mediated transient expression of histidine-tagged CBSV coat protein in *Nicotiana Benthamiana* using pEAQ-HT vector. Protein expression was monitored by Western blot using anti-histidine antibody. The expressed protein was purified using nickel affinity chromatography and identified by mass spectometry. Infiltrated *N. Benthamiana* leaves progressively produced CBSV coat proteins from 3dpi to 9dpi, after which expression declined. The purified 45 kDa protein was identified by 60% sequence coverage as CBSV coat protein. This protein will be used for polyclonal antibody production for the optimization of simple, cost effective CBSV detection tests suitable for CBSV monitoring in West Africa.
Keywords: Cassava, *Cassava brown streak virus*, Transient protein expression

P1164: Other Plant Species

Phenotypic Consequences of Silencing the Primary Nocturnal Carboxylase and Its Circadian Phospho-Regulator in the Crassulacean Acid Metabolism Model K*alanchoë*

Susanna Boxall, University of Liverpool, Liverpool, United Kingdom

Species of the Madagascan succulent genus Kalanchoë have adapted to growing in a semi-arid, sub-tropical environment in part through evolving a modified pathway of photosynthetic CO2 fixation known as Crassulacean acid metabolism (CAM). CAM species increase water use efficiency by opening their stomata for primary CO2 fixation in the dark, when it is cooler and more humid, and closing their stomata in the light, when the atmosphere is at its hottest and driest. The circadian clock underpins optimised temporal regulation of CAM, preventing futile cycling between conflicting metabolic steps. Primary nocturnal CO2 fixation is catalyzed by phosphoenolpyruvate carboxylase (PPC). Temporal control of PPC is achieved via protein phosphorylation, which is catalysed by a specific circadian clock controlled protein kinase, named PPCK1. In the Hartwell lab, our main goal is to identify and characterise functionally all of the genes required for the establishment and regulation of CAM. To facilitate this, the genomes and CAM-associated transcriptomes of several members of the genus Kalanchoë have been decoded and catalogued. Candidate CAM genes are being tested by generating RNAi and/ or constitutive over-expressor transgenic lines. Examples will be presented to highlight our recent progress towards defining functionally each key component of the CAM genetic blueprint. In particular, recently published and unpublished phenotypic data for transgenic Kalanchoë lines lacking the CAM-specific PPC and PPCK1 isogenes will be presented. These lines have revealed unexpected changes in the central circadian clock and thus provide the first evidence for feedback communication between CAM-associated metabolites and the core oscillator.

P1165: Other Plant Species

Transcriptome and Metabolome of *Ferula asafoetida* an Endangered Medicinal Plant

Hajar Amini[1], Mohammad R. Naghavi[2], Shen Tong[3], Yanhong Wang[4], Ikhlas Khan[4], Oliver Fiehn[3] and Julin N Maloof[1], (1)University of California, Davis, CA, (2)Tehran University, Tehran, Iran (Islamic Republic of), (3)NIH West Coast Metabolomics Center, University of California, Davis, CA, (4)National Center for Natural Products Research, University of Mississippi, Oxford, MS

Ferula asafetida, an endangered medicinal plant, is well known as an important source of oleo-gum-resins. These resins possess antifungal, anti-diabetic, anti-inflammatory, anti-mutagenic, and antiviral activities, and are therefore useful for therapeutic industries. In spite of its medicinal attributes, most of the compounds and enzymes involved in secondary metabolite biosynthesis remain uncharacterized at the molecular level. We decided to evaluate transcriptome and metabolome of different tissues of *F. asafoetida* to identify candidate pathways involved in synthesis of medicinal metabolites. Since the major medicinal metabolite components are present in the roots of *F. asafoeida* our study is focused on the comparative *de novo* transcriptome analysis of root versus leaf, stem and flower tissue to identify root specific transcripts involved in terpenoids, sesquiterpens and sesquiterpene coumarins biosynthesis. Because TLC analysis revealed the presence of some sesquiterpene coumarins such as umbelliprenin in methanolic extracts of the root and leaf; and umbelliferone in root and stem we decided to quantify them. Metabolome analysis is being done to quantify umbelliprenin with LC-ESI-MS; these results will be compared to mRNA levels defined by RNA sequencing for the different tissues (root, stem, lead and flower) to identify candidate biosynthesis genes for umbelliprenin.

P1166: Other Plant Species

The Complete Genome Sequence of the Parasitic Weed *Orobanche cumana* (Sunflower Broomrape)

Nicolas Pouilly[1], Jérôme Gouzy[1], Marie-Claude Boniface[1], Olivier Bouchez[2], Sébastien Carrère[1], Olivier Catrice[1], Stephane Cauet[3], Clotilde Claudel[4], Ludovic Cottret[1], Sebastien Faure[4], Álvaro Calderón González[5], Xavier Grand[1], Luyang Hu[1], Céline Jéziorski[3], Marc-Marie Lechat[6], Ludovic Legrand[1], Johann Louarn[1], William Marande[7], Nicolas Ribière[4], Erika Sallet[1], Philippe Simier[6], Leonardo Velasco[5], Cécile Donnadieu[2], Christophe Jestin[8], Philippe Delavault[6], Hélène Bergès[3], Marie Coque[4], Begoña Pérez-Vich[5] and Stéphane Munos[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (2)INRA GeT Genomics Facility, Castanet-Tolosan, France, (3)INRA - CNRGV, Castanet-Tolosan, France, (4)BIOGEMMA, Mondonville, France, (5)Instituto de Agricultura Sostenible - CSIC, Córdoba, Spain, (6)Laboratoire de Biologie et Pathologie Végétales, SFR 4207 QUASAV, Université de Nantes, Nantes, France, (7)French Plant Genomic Center CNRGV - INRA, Castanet-Tolosan, France, (8)Terres Inovia, Thiverval-Grignon, France

Orobanche cumana Wallr. is an obligatory and non-photosynthetic root parasitic plant of the sunflower crop, causing important yield losses on infested fields. Located in Earthen Europe, Spain and Asia, this parasitic weed can rapidly spread to new areas and his emergence has been observed in France since 2007. In sunflower, breeding for resistance was mainly based on single major resistance genes. New more virulent races of *O. cumana* appeared, leading to a breakdown of resistance genes. A better understanding of the mechanisms involved in the interaction

between sunflower and *O. cumana* may improve sustainability of the resistance by using resistance genes acting at different steps of the life cycle of the parasitic plant. Combining long read sequencing (PacBio SMRT technology), optical mapping (BioNano Irys system), SNP-based genetic mapping and RNA-seq expression analysis, we have produced a first version of the 1.42 Gb genome sequence of *O. cumana* (2n=38). Our *de novo* strategy resulted in an assembly of 1.38 Gb, constituted by 622 scaffolds with a N50 of 5.9 Mb, and is provided to the public research community through a Web Genome Browser. We aim to obtain the sequences of the pseudomolecules through an improved genetic map thanks to polymorphism located on the scaffolds, identified by whole genome re-sequencing (Illumina plateform) of the parental lines of a F2 segregating population. The genome sequence of *O. cumana* will contribute to the characterization of its physiology and development and in the understanding of the host-parasite interactions. This release should allow identifying avirulence genes, as putative interactor with sunflower proteins and considering the identification of new resistance genes in sunflower.

P1167: Other Plant Species

A Genome Sequence of Korean Endemic Willow, *Salix koriyanagi* Kimura

Juhyeon Kim, Yongsung Kim and **Jongsun Park**, InfoBoss Co., Ltd., Seoul, South Korea

The genus *Salix* belonging to family Salicaceae contains around 500 species all over the world. Most of Salix species are distributed in Northern hemisphere. 22 Salix species have been identified in South Korea. *Salix koriyanagi* Kimura is Korean endemic willow usually located nearby streams and it has been utilized as a traditional medicine because it contains salicylic and tannic acids. Interestingly, morphology of *Salix purpurea*, of which genome was published as a second genome in *Salix* genus, is very similar to that of *S. koryanagi* and both species were clustered into one clade in molecular phylogenic study. To understand more about *S. koryanagi* genome, we sequenced it using HiSeq 4000 with four different libraries. Total length of genome is 511.97 Mb (N50 is 81,724bp), covering 80% of k-mer analysis result (456 Mb). and maximum length of scaffold is 261,948bp. InterProScan was executed for annotating functional domains from both species, presenting that 85.95% (5,826) functional domains are shared with each other. Around 49Mb sequences of *S. koryanagi* genome were successfully aligned against 65Mb region of *S. purpurea* genome, which is different from similarity of morphology and molecular phylogeny. With in-depth analysis of S. koryanagi genome, relationship between *S. koryanagi* and *S. purpurea* will be established at various levels including genome-wide level.

P1168: Other Plant Species

A Systems Approach to Breeding Disease Resistance in Lettuce Against Necrotrophic Fungal Pathogens

Adam Talbot, University of York, York, United Kingdom

Botrytis cinerea and *Sclerotinia sclerotiorum* are fungal pathogens of global importance, each causing multi-million pound annual crop losses pre- and post-harvest on many dicotyledonous crops including field-grown and protected lettuce crops. Chemical control is and problematic with increased regulation and the emergence of resistance within the microbial population. Development of host resistance is a more sustainable solution, but has been an intransigent problem for breeders.

We are taking a novel approach to breeding for disease resistance against *B. cinerea* and *S. sclerotiorum* in lettuce, combining systems biology and quantitative genetics. Genetic variation in susceptibility to these pathogens was identified in a set of diverse lettuce accessions, based on lesion size following leaf inoculation. Susceptibility to both pathogens was also correlated across different accessions thereby increasing the potential of identifying alleles conferring broad resistance. The validity of the detached leaf assay data was investigated in polytunnel trials, which demonstrated correlation between the susceptibility levels of detached leaves and infected lettuce plants.

To identify genetic loci that contribute to disease resistance, a quantitative genetics approach was employed. Recombinant inbred lines (RILs) from a mapping population were assessed for disease susceptibility using the detached leaf assay and QTLs indicating potential genetic loci contributing to resistance were identified. Further work is being initiated to fine map and validate these QTLs.

We have developed a network analysis gene disco very strategy in Arabidopsis exploiting time series transcriptome data and are now applying this methodology in lettuce to predict genes conferring disease resistance against *B. cinerea*. Using RNAseq we generated a time series of gene expression from 9 to 54 hours post infection in lettuce leaves after inoculation with *B. cinerea* or mock inoculation. From this time series data we will investigate the chronology of transcriptional reprogramming in lettuce following infection, infer regulatory networks mediating this response and predict key regulators of disease resistance within these networks. The expression of these key genes will be tested in diversity set lines showing extreme resistance phenotypes to examine correlation of expression with resistance. Co-localisation of these key regulators with resistance QTL would fast-track the identification of causal genes and associated markers for integration into lettuce breeding programmes.

P1169: Other Plant Species

Effector-Driven Identification of Downy Mildew Resistance Genes in Lettuce

Kelsey Wood, Kyle Fletcher, Juliana Gil, Ayumi Gothberg, Maria Jose Truco, S. Reyes Chin-Wo, Joan Wong and Richard Michelmore, Genome Center, University of California, Davis, Davis, CA

Identification of effectors in plant pathogen genomes is of major interest for understanding the mechanisms of pathogenesis, for monitoring field pathogen populations, and for breeding pathogen resistant plants. Using comparative genomics and bioinformatics, we have identified candidate effectors from the economically important downy mildew species *Bremia lactucae* by searching for the WY domain, a conserved structural element found in *Phytophthora* effectors that has been implicated in their immune-suppressing function. Searching for the WY-domain uncovered additional effector candidates that were missed by searching using traditional effector prediction techniques. The candidate effectors show several characteristics of pathogen effectors, including an N-terminal secretion signal, lineage specificity, and evidence of gene duplication and gene family expansion. Functional characterization of seven of the WY domain effectors revealed three effectors that elicited an immune response on lettuce containing introgressions from wild lettuce species. One of the lettuce varieties showing recognition was used to generate a mapping population and genotyping-by-sequencing was used to map the recognition to a region on chromosome 4 that co-segregates with several known downy mildew resistance genes. This research highlights the usefulness of an effector-driven approach for identifying new resistance genes and for breeding disease resistant crops.

P1170: Other Plant Species

Genetic Analysis of Important Traits in Lettuce

Lei Zhang, Wenqing Su, Chenhuan Yan, Rong Tao, Weiyi Zhang, Yue Jia and **Hanhui Kuang**, Huazhong Agricultural University, Wuhan, China

Lettuce is one of the most important vegetable crops worldwide. As a leafy vegetable, the color and shape of leaves are important horticultural traits of lettuce. We used biparental segregating population and natural population to genetically dissect these traits in lettuce. GWAS analysis identified at least five loci controlling leaf color of lettuce, which were verified using biparental segregating populations. Furthermore, dozens of eQTLs for genes in the anthocyanin biosynthesis pathway were identified using GWAS. Pyramiding of these loci a cultivar is expected to increase anthocyanin concentration in leaves. Leaf color is also affected by the concentration of chlorophyll, and natural mutations in the *GLK* and *PhyB* genes contribute considerably variation in green intensity among different lettuce cultivars. The natural variation of the *PhyB* gene also changes leaf angle and flowering time in lettuce. As many as five loci controlling heading formation in crisphead were identified using a combination of bulk segregant analysis (BSA) and second generation sequencing. Fine mapping showed that knockout mutation of the *STM* gene is necessary but not sufficient for the heading in lettuce. The *STM* gene is also involved in undulation of leaf margin. The underlying molecular mechanisms of these genes are being studied. Our results provide critical information for molecular design breeding of lettuce cultivars.

P1171: Other Plant Species

Development of Lettuce (*Lactuca sativa*) Lines with High Levels of Flavonoids

Csanad Gurdon, Shukhratdzhon Satorov, Meg Tsai, Alexander Poulev and Ilya Raskin, Rutgers University, Department of Plant Biology, New Brunswick, NJ

Flavonoids and phenolic acids have been demonstrated to have a beneficial effect against diabetes and obesity. We have previously developed red lettuce cultivars with high flavonoid and phenolic acid content, and demonstrated their anti-diabetic effect both *in vitro* and *in vivo*. As different flavonoids are expected to have specific health benefits, we aimed to develop further lettuce varieties with an altered flavonoid profile. We report the phytochemical and genetic characterization of three such lettuce mutants: two with high levels of kaempferol, and one with high levels of naringenin chalcone. They were identified in a screen for altered color in a segregating population of mutagenized *cv.* Firecracker red lettuce plants. Green color indicated loss of the red anthocyanin cyanidin-3-malonlyglucoside, and accumulation of flavonoid precursors instead. *Kaempferol+ A* and *Kaempferol+ B* mutants contained high amounts of kaempferol, a compound with anti-diabetic, anti-obesity and anti-inflammatory effects that is usually present in low or non-detectable levels in lettuce. Null mutations in the *flavonoid-3' hydroxylase* (*F3'H*) gene were identified in both of these mutant lines. The *Naringenin Chalcone+* mutant had high amounts of naringenin chalcone, a compound reported to have anti-inflammatory effects. The phenotype results from a null mutation in the *chalcone isomerase* (*CHI*) gene. Phenotype segregation ratios of self-pollinated mutants and their siblings revealed that the mutations were inherited as Mendelian recessive loci. *F3'H* mutants were found to be similar to wild type lettuces apart from their color; *CHI* mutants showed pleiotropic effects including an altered growth habit.

P1172: Other Plant Species

Introgression of Alleles from Non-Domesticated Germplasm to Improve Nitrogen Use Efficiency in Lettuce.

David W. Still, California State University Agricultural Research Institute, Pomona, CA and Youngsook You, Cal Poly Pomona, Pomona, CA

California leads the nation in lettuce production, but global warming and environmental sustainability are challenging factors affecting the industry. Lettuce requires frequent irrigation of high quality water and high amounts of nitrogen (N) fertilizers to make yield and size. Water has become a sometimes scarce and an increasingly unpredictable commodity in California. Nitrogen fertilizers can lead to groundwater contamination and are the primary source of nitrous oxide, a potent long-lived greenhouse gas. To decrease the environmental impact of growing lettuce, we are working toward creating genetic tools and germplasm with improved water and nitrogen use efficiency. Diverse germplasm including non-domesticated *Lactuca* species and commercial cultivars of lettuce were screened under non-limiting N and limited N to identify genotypes with the highest N uptake. Under non-limiting N fertilizers, commercial cultivars had high N uptake. Under limited N, unimproved non-domesticated genotypes had the highest N uptake. However, there was almost no correlation in N uptake under the two N levels indicating different mechanisms involved. Using the genotypes with the highest N uptake under limited N, most of which were non-domesticated accessions, crosses were made to introgress alleles into elite germplasm. Several F2 populations were created and are being evaluated for N uptake and subjected to genotyping-by-sequencing (GBS) to aid in selection and mapping. In addition, we have used GBS to identify multiple QTL that are associated have nitrogen and water use efficiency.

P1173: Other Plant Species

Improving the Post-Harvest Quality of Baby Leaf Lettuce

Annabelle Damerum, University of California, Davis, Davis, CA

Lettuce is one of the most widely consumed vegetables, but contains relatively low concentrations of health-promoting phytonutrients, thought to afford protection against numerous chronic diseases. The short shelf life of popular ready-to-eat baby leaf salad mixes translates into vast amounts of waste and is largely influenced by damages incurred to leaves during processing. Nutrient density and shelf life have been measured to vary depending on leaf type and understanding the genetic basis of this variation is important for breeding efforts to enhance post-harvest quality. In this study, lettuce recombinant inbred lines, generated from a cross between wild (*Lactuca serriola*) and cultivated lettuce (*L. sativa*), were analysed for nutrient quality and for physiological, biochemical and biophysical leaf traits associated with cell wall strength, identifying over 200 quantitative trait loci (QTL) from four environments. QTL "hotspots", where QTL for the same traits measured in different environments co-located to the same genomic region, were mined for candidate genes for nutritional quality and shelf life. Results have been used to inform a commercial breeding programme, to develop novel breeding material from wild lettuce with an improved nutrient quality and ongoing work is utilising CRISPR-Cas genome editing to functionally validate candidate genes for shelf life.

P1174: Other Plant Species

Developing Baby Leaf Spinach with Lower Cadmium Uptake

Juliana Osorio[1], Rachel Greenhut[1], E. Charles Brummer[1], Allen Van Deynze[2], Richard Smith[3] and Timothy K Hartz[4], (1)University of California, Davis, Davis, CA, (2)University of California, Davis, CA, (3)UC Cooperative Extension, Salinas, CA, (4)UC Cooperative Extension, Davis, CA

Spinach (*Spinacia oleracea*) is an important crop in the California Salinas Valley and a known hyper-accumulator of cadmium (Cd). The valley has areas with anomalously high Cd in bedrock and agricultural soils, and recent Food and Drug Administration findings of trace amounts of Cd in baby-leaf spinach led to voluntary product recalls. No current regulation limits Cd content in domestic foods, however the spinach industry is motivated by the potential health risks and economic loss to eliminate Cd contamination in spinach. Due to the ubiquitous nature of Cd in these soils and the lack of grower tools to manage accumulation in spinach, the industry has turned to plant breeders to develop new spinach varieties with reduced Cd accumulation. To this end, we aimed to, i) screen international spinach germplasm collections for leaf Cd accumulation, ii) identify single nucleotide polymorphisms (SNPs) associated with low Cd accumulation within candidate Cd regulatory genes, iii) genotype germplasm collections, and iv) select accessions to incorporate into our spinach breeding program. We found a 3-fold difference in Cd accumulation across tested accessions and made selections for advanced testing and breeding. We also identified SNPs in three candidate Cd regulatory genes and found significant correlation between some alleles and Cd phenotype. This is the first step towards breeding a low Cd accumulating spinach cultivar.

P1175: Other Plant Species

Phenotypic and Genetic Diversity and Association Study of Mineral Components in Spinach

Jun Qin[1], Ainong Shi[1], Beiquan Mou[2] and Michael A. Grusak[3], (1)University of Arkansas, Fayetteville, AR, (2)USDA-ARS, Salinas, CA, (3)USDA/ARS, Houston, TX

Spinach is a useful source of dietary vitamins and mineral elements. The objective of this research was to conduct a genome-wide association study (GWAS) and to identify SNP markers associated with mineral elements in the USDA-GRIN spinach germplasm collection. A total of 14 mineral elements: boron (B), calcium (Ca), cobalt (Co), copper (Cu), iron (Fe), potassium (K), magnesium (Mg), manganese (Mn), molybdenum (Mo), sodium (Na), nickel (Ni), phosphorus (P), sulfur (S), and zinc (Zn) were evaluated in 292 spinach accessions originally collected from 29 countries. Significant genetic variations were found among the tested genotypes as evidenced by the 2 to 42 times difference in mineral concentrations. A total of 2,402 SNPs identified from genotyping by sequencing (GBS) approach were used for genetic diversity and GWAS. Forty-five SNP markers were identified to be strongly associated with the concentrations of 13 mineral elements except K. Three SNP markers, AYZV02017731_40, AYZV02094133_57, and AYZV02281036_185 were identified to be associated with concentrations of four mineral components, Co, Mn, S, and Zn. There is a high validating correlation coefficient with $r > 0.7$ among concentrations of the four elements. Thirty-one spinach accessions, which rank in the top three highest concentrations in each of the 14 mineral elements, were identified as potential parents for spinach breeding programs in the future. The 45 SNP markers strongly associated with the concentrations of the 13 mineral elements except K could be used in breeding programs to improve the nutritional quality of spinach through marker-assisted selection (MAS). The 31 spinach accessions with high concentrations of one to several mineral elements can be used as potential parents for spinach breeding programs.

P1176: Other Plant Species

Phenotypic and Genetic Diversity of Spinach USDA Germplasm Accessions

Ainong Shi[1], **Jun Qin**[1], Beiquan Mou[2] and Jim Correll[1], (1)University of Arkansas, Fayetteville, AR, (2)USDA-ARS, Salinas, CA

Spinach (*Spinacia oleracea* L., 2n=2x=12) is an economically important vegetable crop worldwide and one of the healthiest vegetables due to its high concentrations of nutrients and minerals. The objective of this research was to conduct genetic diversity and population structure analysis of a collection of world-wide spinach genotypes using single nucleotide polymorphisms (SNP) markers. Genotyping by sequencing (GBS) was used to discover SNPs in spinach genotypes. Three sets of spinach genotypes were used: 1) 268 USDA GRIN spinach germplasm accessions originally collected from 30 countries; 2) 45 commercial spinach F1 hybrids from three countries; and 3) 30 US Arkansas spinach breeding lines. The results from this study indicated that there was genetic diversity existed among the 343 spinach genotypes tested and the genetic background in improved commercial F1 hybrids and in Arkansas cultivars/lines and have different structured populations from the USDA germplasm. In addition, the genetic diversity and population structures were associated with geography and germplasm from the US Arkansas breeding program had a unique genetic background. These data could provide genetic diversity information for selecting parents in spinach breeding programs.

P1177: Other Plant Species

Betaine Aldehyde Dehydrogenase (BADH) Expression and Betaine Production in Sugarbeet Cultivars with Different

GuoLong Li, Inner Mongolia Agricultural University, hohhot, China

Many plants accumulate betaine in response to water stress. The betaine aldehyde dehydrogenase (BADH) is the key enzyme in betaine biosynthesis in plants and can be regulated by BADH gene. A pot experiment was conducted to compare the relation of betaine content and BADH gene expression in two sugarbeet (Beta vulgaris L.) varieties differing in drought tolerance. We observed an evident correlation between the transcript up-regulation and the betaine accumulation under water stress. Results showed that the betaine content and BADH gene expression in the drought tolerant variety, was significantly increased under drought stress. The betaine accumulation and BADH gene expression might function as one response to eliminate the negative effects of drought and may play a critical role in maintaining photosynthetic activity in sugarbeet plants under water stress.

Keywords: *Sugarbeet, BADH, Betaine, Water stress*

P1178: Other Plant Species

Transcriptomic Analyses of Taproot Growth and Sucrose Accumulation at Different Developmental Stages in Sugar Beet (*Beta vulgaris* L.)

ShaoYing Zhang, Inner Mongolia Agricultural University, hohhot, China

In sugar beet (*Beta vulgaris* L.), taproot size and sucrose content are important determinants of yield and quality. However, high yield and low sucrose content are two tightly bound agronomical traits. The advances in next-generation sequencing technology and publication of sugar beet genome have provided a method for the study of molecular mechanism underlying the regulation of taproot growth and sucrose accumulation. In this work, we performed comparative transcriptome analyses of the high yield cultivar SD13829 (SD) and the high sucrose content cultivar 04BS02 (BS) at five developmental stages. More than 50,000,000 pair-end clean reads for each library were generated. When taproot turned into the rapid growth stage of 82 day after emergence (DAE), three enriched GO terms, “cell wall”, “cytoskeleton” and “enzyme linked receptor protein signaling pathway”, occurred in both SD and BS. Differentially expressed genes (DEGs) of paired comparison for BS and SD were enriched in “cell wall”. For pathway enrichment analyses of DEGs that were respectively generated at 82 DAE compared to 59 DAE (before taproot turning into the rapid growth stage) in both SD and BS, “plant hormone signal transduction pathway” was enriched. Several transcription factor family members were up-regulated in rapidly increased taproot. An antagonistic expression of brassinosteroid- and auxin-related genes in taproots was necessary for the rapid enlargement of taproots at stage 82 DAE. In SD, the growth strategy was relatively focused on cell enlargement promoted by brassinosteroid signaling, whereas in BS was relatively focused on secondarily cambial cell division regulated by cytokinin, auxin and brassinosteroid signaling. Taken together, our data demonstrate that the size and sucrose content of taproots rely on the taproot growth strategy, which is controlled by cytokinin, auxin, GA and brassinosteroid signaling.

Key words: *sugar beet, taproot growth, sucrose content, RNA-Seq*

P1179: Other Plant Species

Identifying Phenotypes, Markers, and Genes in Carrot Germplasm to Deliver Improved Carrots to Growers and Consumers

Phil Simon[1], Shelby Ellison[1], Douglas Senalik[2], Micaela Colley[3], Julie Dawson[4], Sherry Tanumihardjo[4], David Spooner[5], Edgar Spalding[6], Joe Nunez[7], Allen Van Deynze[8], Dan Sumner[8], Philip A. Roberts[9], Lindsey du Toit[10], Tim Waters[11], Massimo Iorizzo[12], William Matthews[8], Hyunok Lee[8] and Laurie McKenzie[3], (1)USDA, ARS and University of Wisconsin, Madison, Madison, WI, (2)USDA, ARS, Madison, WI, (3)Organic Seed Alliance, Port Townsend, WA, (4)University of Wisconsin, Madison, WI, (5)USDA - ARS, Madison, WI, (6)University of Wisconsin-Madison, Madison, WI, (7)University of California, Bakersfield, CA, (8)University of California, Davis, CA, (9)Department of Nematology, University of California Riverside, Riverside, CA, (10)Washington State University, Mount Vernon, WA, (11)Washington State University, Pasco, WA, (12)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC

A survey of U.S. carrot growers and seed industry stakeholders was conducted and a meeting was held in 2015 to identify key traits important for improved carrot quality and productivity anticipated to meet future market demands. This effort revealed that the carrot industry needs breeding stocks and genomic tools that can be used to develop carrots with improved field performance including disease and pest resistance, and abiotic stress tolerance; and improved flavor and nutritional quality to better meet consumer needs. Given this critical stakeholder input, the goals of this project are to: 1) phenotype diverse carrot germplasm and breeding stocks to discover and characterize previously uncharacterized variation for traits important for improving carrots for the US market; 2) develop an expanded carrot genomic and phenotypic database for breeders to catalogue genomic variation and track genes underlying important traits; 3) initiate the development of breeding pools from diverse germplasm and breeding stocks that include alleles for improved crop production and consumer quality traits, and test them on-farm with growers and for flavor and nutritional value for consumers; and 4) evaluate the market value and impact of carrot traits on grower and consumer decisions. A timeline of activities for this project has been developed.

P1180: Other Plant Species

Mapping Black Spot Disease Resistance and Molecular Maker Development in Tetraploid Roses.

Cindy Rouet[1], Daryl J. Somers[1], Travis W. Banks[1], Parminder Sandhu[1] and Elizabeth Lee[2], (1)Vineland Research and Innovation Centre, Vineland Station, ON, Canada, (2)University of Guelph Dept. of Plant Agriculture, Guelph, ON, Canada

Black spot is the most deadly foliar disease of landscape roses. It reduces ornamental value and marketability of the plants. Genetic resistance to black spot is available but is difficult to breed for using visual ratings to evaluate hybrids in the field. Therefore, black spot disease resistance is a good candidate for developing marker-assisted selection in roses. A high-density, integrated SNP map of all 56 rose chromosomes was created using GBS on 300 progeny in a tetraploid biparental cross ‘564’ x ‘Gentle Giant’. The map contained 2,281 SNPs with a length of 821 cM. Resistance to black spot race 10 was mapped and markers were identified in this population. Our current research has created a second SNP map in 350 progeny of ‘CA60’ x ‘SITR’ where we again mapped black spot race 10 resistance and a newly identified race (isolate VSK04). Phenotyping used an improved detached leaf assay and both race-specific resistance genes mapped to the same location. Research is on-going to generate a consensus map in order to identify markers associated with black spot resistance that would be widely applicable in rose breeding programs. We also have preliminary evidence that genotype ‘564’ carries horizontal/quantitative resistance to black spot disease (isolate VSK04) which differs from the qualitative resistance mapped previously. A whole genome sequence of diploid *Rosa multiflora* will be generated soon to assist the rose breeding program with developing markers that are diagnostic across multiple breeding populations.

P1181: Other Plant Species

Mapping a New Black Spot Resistance Locus in Rose

Jason D. Zurn[1], David C. Zlesak[2], James Bradeen[3], Stan C Hokanson[4] and Nahla Bassil[1], (1)USDA-ARS National Clonal Germplasm Repository, Corvallis, OR, (2)University of Wisconsin River Falls, Department of Plant and Earth Science, River Falls, WI, (3)University of Minnesota, Department of Plant Pathology, St. Paul, MN, (4)University of Minnesota, Department of Horticultural Science, St. Paul, MN

Rose black spot, caused by *Diplocarpon rosae*, is one of the most devastating foliar diseases of cultivated roses (*Rosa hybrida*). The pathogen is globally distributed and has the potential to cause large economic losses in the outdoor rose industry. Genetic resistance is the most economical disease management strategy for black spot and many breeding programs are focused on creating cultivars with durable resistance. The tetraploid cultivar Brite Eyes™ ('RADbrite') is resistant all *D. rosae* races except race 12. Because of this broad resistance, a 94 individual F_1 mapping population was developed by crossing Brite Eyes™ to the susceptible tetraploid 'Morden Blush'. The population was phenotyped with four races (8, 9, 10, and 11) and a genetic map was constructed using the WagRhSNP 68K Axiom array. The F_1 individuals were either resistant or susceptible to all races evaluated and segregated 1:1, suggesting resistance is mediated by a single locus. Preliminary mapping places the Brite Eyes™ resistance locus on a single linkage group in a 31.7 cM region delimited by RhMCRND_7069_318 and Rh12GR_258_2610. This linkage group is homeologous to chromosome 5 of the diploid integrated consensus map. Prior to this experiment, three resistance loci have been identified (*Rdr1, Rdr2,* and *Rdr3*). Both *Rdr1* and *Rdr2* are located on a chromosome 1 homeolog and the equivalent chromosomal location of *Rdr3* is unknown. However, races 3 and 9 are virulent on *Rdr3*, suggesting the resistance in Brite Eyes™ is novel. Future efforts will focus on developing a diagnostic test for marker assisted selection.

P1182: Other Plant Species

An Artificial Freezing Test to Screen for Cold Hardiness in *Rosa hybrida*.

Daryl J. Somers, Elodie Derivry, Cindy Rouet and Parminder Sandhu, Vineland Research and Innovation Centre, Vineland Station, ON, Canada

The national rose breeding program in Canada that was initiated in the early 1900's has released numerous cold hardy hybrid rose varieties due to repeated testing and selection in extreme cold environments. Roses in the collections are adapted to USDA climatic zones 2 and 3 and can withstand temperatures as low as -40C. Currently the Canadian rose breeding program is located in Vineland, ON (USDA climatic zone 5a). In order to continue breeding cold hardy roses, the initial hybrid selections from Vineland will need to be pre-screened for cold hardiness to eliminate susceptible genotypes prior to testing in cold western Canadian locations. An artificial screening method was employed whereby potted roses were acclimated at 4C for 14 days prior to programmed freezing experiments. Leaf and stem tissues from acclimated plants were subjected to freezing in a programmable freezer in which the temperature ramped down from 4C to -25C at 2.5C/h. Electrolyte leakage was used to estimate the degree of injury due to freezing of stem sections. The correlation between electrolyte leakage at -15C using artificial freezing of stem sections and 2 years of replicated, multi-location field-based freezing injury data collected in USDA zone 3 (Alberta and Saskatchewan) was 0.82. This strong correlation will permit efficient pre-screening of genotypes prior to lengthy, replicated field testing. We are currently evaluating a biparental cross using artificial freezing where we intend to genetically map cold hardiness in roses using GBS.

P1183: Other Plant Species

Comparative Transcriptome Analysis of Heat Stress Responsiveness in Two Contrasting Ginseng Cultivars

Murukarthick Jayakodi[1], Sang-Choon Lee[1] and Tae-Jin Yang[2], (1)Seoul National University, Seoul, Korea, Republic of (South), (2)Seoul National University, Seoul, South Korea

Panax ginseng (Asian/Korean ginseng) has been used in traditional medicine to strengthen the body and mental well-being of humans for thousands of years. Many elite ginseng cultivars were developed and cultivation was well established during last hundred years. However, heat stress in an important threat limitation to growth and sustainable production of ginseng. Efforts have been made to study the effects of high temperature on ginseng physiology. However, the knowledge of the molecular responses to heat stress is still limited in ginseng. Here, we sequenced the transcriptomes (RNA-Seq) of two ginseng cultivars with different heat sensitivity, Chungpoong (CP) and Yunpoong (YP), which are sensitive and strong cultivars to heat stress, respectively, after 1- and 3-week heat treatments. CP is more sensitive to heat stress than YP resulted in higher chlorophyll contents in YP than that in CP. Moreover, heat stress declined chlorophyll contents more rapidly in CP compared to YP. A total of 329 heat-responsive genes were identified, which showed similar expression pattern against heat stress in both cultivars. Intriguingly, members of chlorophyll ab binding (CAB), WRKY transcription factors and fatty acid desaturase (FAD) gene family were found to be majorly responded during heat stress, which in turn inhibition of photosynthesis. Further, a genome-wide scan of photosynthetic and sugar metabolic genes revealed the implication of reduced transcripts level of ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBISCO) enzyme with elevated soluble sugars. This work was supported by the grant funded by Next-Generation BioGreen21 Program (No. PJ01103001, PJ01100801), Rural Development Administration, Republic of Korea.

P1184: Other Plant Species

Comprehensive Comparative Analysis of Seven Chloroplast Genomes of *Panax* Species and Establishment of Authentication System Based on Species-Unique SNP Markers Derived from Chloroplast Genomes

Binh Van Nguyen[1], Giang Ngoc Linh Vo[1], Nomar Espinosa Waminal[2], Hyun-Seung Park[1], Jee Young Park[1] and Tae-Jin Yang[1], (1)Seoul National University, Seoul, South Korea, (2)Seoul National University, Seoul, Korea, Republic of (South)

Panax species belongs to the Araliaceae family and regarded as the king of herbal medicinal plants. However, their evolutionary, taxonomical relationship, and origin remain largely unresolved. In recently we reported complete chloroplast genomes and 45S nuclear ribosome DNA (nrDNA) sequences from seven *Panax* species. Two species, *P quinquifolius* and *P. trifolius*, are derived from North America and five species, *P. ginseng*, *P. notoginseng*, *P. japonicus*, *P. vietnamensis*, and *P. stipuleanatus* are derived from Asia. Comparative analysis of chloroplast sequences of seven *Panax* elucidate the genetic diversity, evolution relationship and develop markers for their authentication. Complete chloroplast genomes of seven *Panax* species are ranged 155,993 ~ 156,466 bp and consisted of 113 functional genes without structural variation between species. Phylogenetic analysis revealed that the nineteen Araliaceae species could be divided into two monophyletic lineages consisting of the *Aralia-Panax* group and the other group comprising the seven remaining genera. Numbers of SNP and repetitive sequences were identified in 7 *Panax* chloroplast genomes including 1,128 SNPs in CDS regions and 1,783 SNPs, 319 repetitive sequences in whole chloroplast genomes. We compared chloroplast genome and mitochondrial genomes and found twelve large chloroplast genome fragments were inserted in mitochondrial genome. We developed 18 chloroplast gene-derived SNP markers to authenticate the 7 *Panax* species with two or three unique species-specific markers for each species. These markers were successful to discriminate one species from others, and can apply for protection of industry of each *Panax* species from economically motivated adulteration.

P1185: Other Plant Species

Targeted Capture of Dreb Subfamily Genes As Candidates Genes for Drought Tolerance Polymorphism in Natural Population of Coffea Canephora

Sinara Aquino[1], Pierre Marraccini[2], Cedric Mariac[1], Marie Couderc[1], Kevin Bethune[1], Alan C. Andrade[3], Darracq Olivier[4], Maud Lepelley[5], Yves P. Vigouroux[1], Catherine Kiwuka[6], Niels Anten[7], **Alexandre de Kochko**[8] and Valérie Poncet[1], (1)IRD, Montpellier, France, (2)CIRAD UMR AGAP, Montpellier, Brazil, (3)Embrapa Café/INOVACAFÉ - UFLA, Lavras-MG, Brazil, (4)Nestlé R&D, Tours, France, (5)Centre R&D Nestlé Tours, Tours, France, (6)NARO, Entebbe, Uganda, (7)Wageningen University, Wageningen, Netherlands, (8)IRD UMR DIADE, Montpellier, France

Coffea canephora, (Robusta), provides 33% of worldwide coffee production, 80% and 22% of Ugandan and Brazilian coffee production, respectively. Abiotic stress such as temperature variations or drought periods, aggravated by climate changes, are factors that affect this production. This sensitivity threatens both the steady supply of quality coffees and the livelihood of millions of people producing coffee. The natural genetic diversity of *C. canephora* offer a potential for detecting new genetic variants related to drought adaptation. In particular, modifications occurring in genes related to abiotic stress tolerance make these genes candidate for breeding programs in order to enhance the resilience to climate change.

1. *canephora* transcription factors from the *DREB* subfamily (Dehydration Responsive Element Binding Protein) have been recently identified as candidate genes. Indeed, in the *C. canephora* Conilon group, the *CcDREB1D* gene showed an increased expression in response to drought in the leaves of a drought tolerant clone[1].

The objectives of this study are to identify and characterize the allelic diversity (Single Nucleotide Polymorphism, SNPs) within drought-tolerant candidate genes with a special focus on the DREB subfamily genes. These genes will be annotated on the reference genome sequence of *C. canephora*[2,3] and on a new assembly. A targeted capture array[4] will be designed for these entire genes, and their flanking regions. These captured regions will be sequenced in a set of wild Ugandan populations. Subsequent detection of SNPs for the whole set will be used to test correlation of these SNPs with traits related to drought tolerance.

We expect to understand the adaptive strategies developed by crops in the wild in order to respond to climate change and use the genetic resources within wild populations, as a basis for transferring drought- and heat-tolerance traits.

[1]Marraccini, P. et al. (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *Journal of experimental botany*, **63**, 4191-212.

[2]Denoeud F. et al. (2014) The coffee genome provides insight into convergent evolution of caffeine biosynthesis. Science 345(6201): 1181

[3]Dereeper, A. et al. (2015) The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Res*, **43**, D1028-35.

[4]Mariac C. et al. (2014). Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. Mol Ecol Resour 14: 1103-1113

P1186: Other Plant Species

Better Coffee Quality from the Lower Canopy?

Bing Cheng, QAAFI, The University of Queensland, St Lucia, Queensland, Australia

From an evolutionary viewpoint, seeds have developed to store nutrients to support plant reproduction. They also have a defence system to survive environmental treats to survival of the seed. With the threat of climate change, a comprehensive understanding of how the growth environment influences seed maturation and composition is required. The Arabica coffee bean, a tropical dicotyledonous albuminous seed, was used in this research to study how seed ripening was influenced by the micro-environment as determined by canopy position. The transcriptome of the developing coffee beans (covered by green, yellow and red pericarp) was analysed for both the upper and the lower canopy (above and below 170 cm). A long read coffee transcriptome obtained from the same samples was used as a reference. Comparative transcriptome analysed was used to investigate the influence of canopy position at different developmental stages. Phenotypic variations of the bean were analysed, including the key physical and chemical attributes influencing coffee quality. Additional sensory testing was also conducted to investigate the aromatic differences in the beans. This comprehensive study will facilitate an improved understanding of the molecular basis coffee quality and provide insights to the variation of seed ripening in different canopy position that produce different phenotypic traits.

P1187: Other Plant Species

First Insights into the Genetic Factors Underlying Post-Flowering Drought Tolerance in Flax

Braulio J. Soto-Cerda[1], Rocío Quian[1], Humberto Gajardo[1] and Sylvie Cloutier[2], (1)Centro de Genómica Nutricional Agroacuicola, Temuco, Chile, (2)Agriculture and Agri-Food Canada, Ottawa, ON, Canada

Drought stress acts simultaneously on many traits and developmental stages which ultimately affects seed yield of crops worldwide. In particular, drought stress at the reproductive stage (flowering and seed development) can result in the most significant reductions in crop production. In this study, yield and yield-related drought tolerance indices were assessed on 120 flaxseed accessions under mild drought and irrigated conditions at flowering time. A set of ~700,000 single nucleotide polymorphisms (SNPs) was screened for marker-trait associations using general models. In average, drought stress reduced seed yield by 18% and start of flowering was brought forward by 5 days. Based on the stress tolerance index (STI), three genomic regions on linkage groups 2, 7 and 9 were identified which explained 68.2% (R^2) of the phenotypic variation. Ten candidate genes were located nearby peak SNPs (~50 kb either side) where a clathrin heavy chain 1 (CHC1), a mitogen-activated protein kinase (MAPK), a peptide chain release factor and a CCCH-type zinc finger protein genes, which have previously been involved in drought response in other crops were the most promising candidates for further studies. This information provides important genetic insights into the natural variation of flaxseed drought tolerance. The identified SNPs or candidate genes could serve as direct targets for both genetic engineering and selection for flaxseed trait improvement.

This research was supported by Fondo Nacional para el Desarrollo Científico y Tecnológico (FONDECYT) Chile, project N° 1161133.

P1188: Other Plant Species

The Flax Genome – a Revolution in Evolution through Natural Genome Editing

Christopher Cullis, Case Western Reserve University, Cleveland, OH

The flax genome can be rapidly modified within a single generation in response to the growth environment. The sites of variation have been characterized in 5 lines, the progenitor line (Pl), 3 derived lines (genotrophs) and a tissue cultured line, by whole genome sequencing. The variations do not appear to be due to either random mutations or the movement of transposable elements, but rather that the genome appears to be able to be edited resulting in two well-defined, different sequences at many loci. These edited regions can extend over many kilobases. Such large-scale, reproducible variation is unlikely to occur through multiple independent events. Therefore an editing mechanism by which long tracts of the genome can be replaced with an alternative structure is being proposed. This editing mechanism is activated somatically, in the apical meristem, by the sub-optimal growth of an individual, with a set of genomic sites available for modification that have the potential to generate phenotypic diversity without lethality. These sites can be independently modified until a 'new' genome, that functions more efficiently in a meristematic cell the selecting environment, is generated. At that stage the editing mechanism is silenced and the 'new' genome replaces the previous version. If this occurs prior to flowering, the edited genome is transmitted to the progeny. Both the initial and the 'new' genomes have been observed in flax varieties and the wild progenitor of cultivated flax.

P1189: Other Plant Species

Chrysanthemum *CmHSFA4* Gene Positively Regulates Salt Stress Tolerance in Transgenic Chrysanthemum

Sumei Chen and Shuang Zhao, Nanjing Agricultural University, Nanjing, China

Salinity induced Na^+ toxicity and oxidative stress hamper plant growth. Here, we showed that expression of the chrysanthemum *CmHSFA4*, a homologue of the heat shock factor *AtHSFA4a*, is inducible by salt, and localizes to the nucleus. It is a transcription activator binding with HSE. Chrysanthemum overexpressing *CmHSFA4* displayed enhanced salinity tolerance by limiting Na^+ accumulation while maintaining K^+ concentration, which is consistent with the up-regulation of ion-transporters *CmSOS1* and *CmHKT2*. Additionally, the transgenic plants reduced H_2O_2 and $O_2^{\cdot-}$ accumulation under salinity, which could be due to up-regulation of ROS-scavenger activities such as SOD, APX and CAT as well as *CmHSP70, CmHSP90*. Together, these results suggest that *CmHSFA4* conferred salinity tolerance in chrysanthemum as a consequence of Na^+/K^+ ion and ROS homeostasis.

P1190: Other Plant Species

Transcriptomic Analysis of Differentially Expressed Genes in the Floral Transition of the Flowering Chrysanthemum

Haibin Wang, Nanjing Agricultural University, Nanjing, China

Chrysanthemum is a leading cut flower species. Most conventional cultivars flower such as 'Jimba' during the fall, but the Chrysanthemum morifolium 'Yuuka' flowers during the summer, and 'QX' during in winter, thereby filling a gap in the market. To date, little is known about molecular basis of flowering time in the flowering chrysanthemum. Here, the genome-wide transcriptome of different cultivars was acquired using RNA-Seq technology, with a view to shedding light on the molecular basis of the shift to reproductive growth as induced by variation in the photoperiod. The transcriptomic analysis and the other data of the spontaneous mutant of chrysanthemum 'Jimba' showed that higher GA contents can affect the expression of SOC1 under LD conditions for regulating flowering; under SD conditions, the photoperiod pathway majorly acted to determine flowering time through CmFTL3, and the GA signaling pathway played a subsidiary role for flowering. In the summer-flowering cultivar 'Yuuka', under short days, genes acting in the photoperiod and gibberellin pathways might accelerate flowering, while under long days, the trehalose-6-phosphate and sugar signaling pathways might be promoted, while the phytochrome B pathway might block flowering. The floral transition of winter-flowering Chrysanthemum is a complex biological process regulated by a wide range of factors, during which time many flowering pathways such as photoperiod pathway, ambient temperature pathway, and vernalization pathways become activated.

P1191: Other Plant Species

Cyanobacteria Is Uniquely Enriched in the Roots of Grain Amaranthus

Dr. Shubham Dixit[1], Suran R Nambisan[1], Suvratha J[1], Bibha Choudhary[2] and Subhashini Srinivasan[3], (1)Institute of Bioinformatics and applied biotechnology, Bangalore, India, (2)Institute of Bioinformatics and Applied Biotechnology, Bangalore,, India, (3)Institute of Bioinformatics and Applied Biotechnology, Bangalore, Karnataka, India

Plants and microbes have co-evolved symbiotically for millions of years driving positive selection. The bartering of chemicals produced by microbes and plants favor enrichment of certain types of bacteria providing advantage to the plant under a given environment. Grain amaranths display certain important agronomic characteristics like C4 dicot, high protein and high lysine grains, resistance to biotic and abiotic stress. Considering an unusual collection of desirable traits by grain amaranths, it is worth exploring the role of bacteria in the evolution of these traits. The objective of this study is to identify bacterial root microflora unique to grain amaranths. Here, we compared rhizospheric and endophytic composition of 16SrRNA using metagenome sequencing of roots from selected species under major plant orders including the three varieties of grain amaranths. We found that Cyanobacteria are uniquely enriched in grain amaranths. This observation is given further credence by the number of reads observed in the root transcriptome of *Amaranthus hypochondriacus*, one of the grain amaranths. We have reconciled the 16SrRNA sequences from culturable Cyanobacteria to those sequenced from roots of one of the grain amaranths and have characterised them biochemically. We have shown that the identified culturable Cyanobacterial species had different level of nitrogen fixation capabilities, as each cyanobacterial species has different level of nitrate reductase enzyme, an important enzyme for nitrogen fixation. We conclude that roots of grain amaranths selectively enrich for Cyanobateria to fix nitrogen.

P1192: Other Plant Species

The Structure of Sex Chromosomes in *Rumex acetosa*

Jan Safar[1], Wojciech Jesionek[2] and Roman Hobza[2], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Institute of Biophysics, Brno, Czech Republic

In contrast to animals, the presence of species with separated sexes with well-established sex chromosomes is very rare in plant kingdom. *Rumex acetosa* is a dioecious plant with XX/XY_1Y_2 chromosome system. Although the evolution of sex chromosomes has been the subject of numerous studies, global view of sex chromosome structure is still missing in this species. We have flow-sorted and sequenced sex chromosomes and autosomes in *R. acetosa*. We focused on the role of various repetitive elements in the process of Y chromosomes formation. Our data demonstrate that *R. acetosa* genome was formed by expansion of various repetitive elements with specific pattern of distribution in case of sex chromosomes. We show that some tandem repeats and retroelements are ubiquitous in the *R. acetosa* genome but surprisingly absent on the Ys or X chromosome. We have quantified the extent of accumulation of individual repeats and they potential role in sex chromosome evolution in *R. acetosa*.

Acknowledgements: The work was supported by the Czech Science Foundation (grant number 16-08698S) and by the award LO1204 from the National Program of Sustainability I.

P1193: Other Plant Species

High-Throughput SNP Discovery and Genotyping in Unique Genetic Resources for Constructing a High Resolution Genetic Linkage Map of *Allium cepa*

Masayoshi Shigyo, Yamaguchi University, Yamaguchi, Japan, Shusei Sato, Tohoku University, Sendai, Japan and Hideki Hirakawa, Tohoku University, Kisarazu, Japan

Shallot (*A. cepa* Aggregatum group) has been applied for several important material resources such as *A. fistulosum*-shallot monosomic addition lines (MALs). With the aim to make the best use of these resources, transcriptome analysis of doubled haploid shallot ('DHA') was carried out by applying Illumina sequence technology. By assembling 25 million qualified reads, the data set of 56,161 bulb unigenes were obtained and used as reference data set. Then the transcriptome sequence reads obtained from each MAL were mapped onto the reference unigene, and the genotyping of identified SNP sites were carried out. By comparing the genotype call of parental *A. fistulosum* and shallot transcriptome mapping data, SNP sites with alternative homozygous call in *A. fistulosum* and reference homozygous call in shallot were selected. By checking the genotype call of these SNP sites in MAL, anchored chromosome was estimated by the additional shallot chromosome in the MAL with heterozygous call. As a result, more than 20,000 unigenes were anchored onto one of 8 chromosomes. Transcriptome based genotyping were also applied for mapping population derived from cross between 'DHA' and doubled haploid of bulb onion ('DHC'). Genotyping of the SNPs detected by mapping of transcriptome reads from 'DHC' were carried out based on the mapping results of transcriptome sequences from mapping population. By using genotype information on 98 lines of mapping population, a high resolution genetic linkage map with 610 genotype blocks composed of around 4,400 unigenes were constructed.

P1194: Other Plant Species

Virus Proliferation during Orchid Flower Development

Hongxing Yang, Shanghai Chenshan Botanical Garden, Shanghai, China

The orchid family (Orchidaceae) is one of the two most species-enrich families of flowering plants. Many plants of this family are of great values with respect to ornamental, medicinal, or other purposes. However, most orchid plants are susceptible to virus infection, causing great production and economical losses. The Cymbidium mosaic virus (CymMV) and the Odontoglossum ringspot virus (ORSV) are the two most common viruses found in orchids, both positive-strand RNA virus. During the transcriptomic study of the floral development in *Phalaenopsis equestris*, which is one of the most widely used model species of the orchid family, we found the proliferation of CymMV was increasingly active during flower development, whereas that of ORSV was not. The increased proliferation CymMV was more evident in the single-lip wild-type plants, compared with the mutant plants with triple-lips. Coincide with the increased proliferation of CymMV, we observed increased instances of gene expression regulations, and alternative splicing events, especially at the stage of open flowers. Gene function enrichment analysis revealed the pathways related to innate immune responses, as well as other biotic and many abiotic stresses were regulated at the open flower stage, and some were more significantly regulated in the triple-lip mutants. Further analysis suggested the replication of the CymMV and ORSV were increasingly active following flower development. We discussed the possible reasons of the seemingly relapsed control over virus proliferation at the late stages of floral developmental, as well as the potential implications for virus detection and future development of virus-resistant plants.

P1195: Other Plant Species

Inference of the Genes Related to Saponin Biosynthesis Pathway of *Gleditsia sinensis* by *de novo* Transcriptome Analysis

Hideki Hirakawa[1,2], Yusuke Kuwahara[1], Daisuke Nakajima[3], Sayaka Shinpo[3], Michimi Nakamura[4], Noriaki Kawano[5], Nobuo Kawahara[5], Mami Yamazaki[6], Kazuki Saito[4] and Hideyuki Suzuki[3], (1)Tohoku University, Sendai, Japan, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)Kazusa DNA Research Institute, Kisarazu, Japan, (4)Chiba University, Chiba, Japan, (5)National Institutes of Biomedical Innovation, Health and Nutrition, Tsukuba, Japan, (6)Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Japan

Gleditsia sinensis is one of the medicinal plants belonging to Fabaceae, which is used for expectorant, emulgent and so on. The plant contains saponins and is traditionally used as soap. With the advance of the sequencing technologies, the genome information of medicinal plants have been accumulated, but the pathways for the biosynthesis of physiologically active substances have been not yet fully understood. In this study, the genome information are obtained by *de novo* transcriptome analysis to clarify the saponin biosynthesis pathway. Total RNAs were extracted from the nine organs (leaves, flowers, and stems etc), and 14.3 Gbp transcript sequences were sequenced by the Illumina platform. Transcript assembly was performed by Trinity, and the ~48,000 unigenes with high quality were constructed. The unigenes were annotated according to the BLAST searches against NCBI NR database and Blast2GO software. The ~32,000 unigenes (66%) were similar to the known genes, and most of them were similar to the genes of pigeon pea and soybean in Fabaceae. The expression of the genes related to the saponin biosynthesis pathway, cytochrome P450, and UDP glucuronosyltransferase (UGT) were compared among the nine organs, and several tens of genes were found to be highly expressed in the seeds. In addition, the profiling of LC-MS (Liquid Chromatography/Mass Spectrometry) data

has been obtained from the pods, and about 10 saponins were found from the base peak. Currently, the genes related to the saponin biosynthesis pathway will be clarified by integrating the information of the annotation, transcriptome, and metabolome data.

P1196: Other Plant Species

Transcriptome-Wide Identification of Transcription Factors Families in *Saussurea lappa*

Vasundhara Thakur[1], Savita Bains[2], Ravneet Kaur[1] and Kashmir Singh[1], (1)Panjab University, Chandigarh, India, (2)panjab university, chandigarh, chandigarh, India

Saussurea lappa C.B Clarke is well recognized medicinal plant for its pharmacological properties such as antifungal, antidiabetic, anti-tumor, antimicrobial, antiulcer and anti-inflammatory properties. The major active constituents isolated from the plant include terpenes, besides other flavonoids, alkaloids and anthraquinones. Costunolide, a sesquiterpene lactone isolated from plant root, largely attributes for its immense pharmacological potential. Despite of its proven medicinal value, the molecular mechanisms governing its metabolic pathways remain less explored. For this study, we identified genes involved in costunolide biosynthesis pathway. To study the regulation of key genes and transcription factors regulating significant metabolic pathways, we conducted high throughput transcriptome sequencing from young plant leaf tissue using Illumina HiSeq 2000 platform, generating 62,039,614 high quality paired end raw reads. Around 114,049 contigs were obtained after the reads were filtered and cleaned. Homology search against PlantTFDB resulted in identification of 19,135 transcripts, which were classified into 58 plant transcription factor families, with MYB, NAC, WRKY, bHLH and MYB-related families regulating metabolites synthesis. Gene ontology terms were assigned to 17,815 transcripts (93.1%) using Blast2GO software, categorizing into biological process (33%), molecular function (46%) and cellular component (21%). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis mapped transcript IDs to annotate them to terpenoid biosynthesis pathways genes. Bioinformatic analysis of co-expression facilitated identification of chief transcription factors co-regulating the terpenoid biosysnthesis pathway genes. Genome walking technique will be used for promoter cloning and further interaction of these promoters with transcription factors will be studied.

KEYWORDS: *Saussurea lappa,* transcription factors, metabolic pathways, costunolide biosynthesis

P1197: Other Plant Species

High-Throughput Sequencing and *de novo* Assembly of *Scorzonera tau-saghyz* (Lipsch. & G.G.Bosse) for Transcriptome Analysis

Vitaly Portnoy[1], Karim Boguspaev[2], Dmitriy Faleev[3], Anastasiya Kapytina[2], Ulykbek Kairov[4], Askhat Molkenov[5], Adi Doron-Faigenboim[6] and Yaakov (Kobi) Tadmor[7], (1)Agricultural Research Organization, ARO, Ramat Yishay, Israel, (2)Institute of Problems of Ecology, Al-Farabi Kazakh National University, Almaty, Kazakhstan, (3)Institute of Problems of Ecology, Al-Farabi Kazakh National University, Almaty, Kazakhstan, Almaty, Kazakhstan, (4)Laboratory of Bioinformatics and Systems Biology, Nazarbayev University, Astana, Kazakhstan, (5)Laboratory of Bioinformatics and Systems Biology, Nazarbayev University, Astana, Kazakhstan, (6)Department of Vegetable and Field Crops, ARO, Rishon LeZion, Israel, (7)Department of Vegetable and Field Crops Crops, Newe Ya'ar Research Center, ARO, Ramat Yishay, Israel

Karim Boguspaev[1], Dmitry Faleev[1], Anastasiya Kapytina[1], Ulykbek Kairov[2], Askhat Molkenov[2], Adi Faigenboim[3], Yaakov Tadmor[4], Vitaly Portnoy[4]

[1]Institute of Problems of Ecology, Al-Farabi Kazakh National University, Almaty, Kazakhstan

[2]Laboratory of Bioinformatics and Systems Biology, Center for Life Sciences, National Laboratory Astana, Nazarbayev University, Astana, Kazakhstan.

[3]Agricultural Research Organization, Department of Vegetable and Field Crops, Rishon LeZion, Israel

[4]Agricultural Research Organization, Department of Vegetable and Field Crops Crops, Newe Ya'ar Research Center, Ramat Yishay, Israel

Natural rubber, produced exclusively from the rubber tree (*Hevea brasiliensis*), exhibits various specific properties that cannot be fully mimicked by synthetic rubber. Alternative sources to natural rubber are of high interest due to biological, economic and political threats to current supplies. Tau-Saghyz (*Scorzonera Tau-Saghyz* Lipsch. & G. G. Bosse), endemic to Kazakhstan, synthesizes high molecular weight natural rubber, thus it could potentially serve as alternative source to natural rubber. Tau-Saghyz was largely cultivated in the former Soviet Union, as a rubber source, between 1931 and 1955, Currently, Tau-saghyz, which can accumulate up to 40% of rubber in the dry roots, is practically of no interest to world science. RNA-seq was conducted on the RNA isolated from different organs of Tau-Saghyz (leaves, stems, roots and flowers), using Illumina Hiseq 2000 sequencing platform. *De novo* transcriptome assembly was conducted using Trinity program, generating ~275M high-quality reads, which were assembled into 203,857 transcripts with an average sequence length of 588 bp and N50=757 bp. Functional annotation was performed using BLASTX homology searches against the NCBI non-redundant (nr) protein and Swiss-Prot databases. Gene sequences and differential expression of enzymes known to be involved in the biosynthesis of natural rubber were analyzed. Our RNA-Seq results analysis identified genes and transcript profiling, in a non-model plant, which could be of high interest for future breeding of Tau-Saghyz and biotechnological approaches.

P1198: Other Plant Species

Transcriptome Plasticity underlying the Plastic Phenotypes of *Alternanthera philoxeroides* in Contrasting Hydrological Conditions

Lexuan Gao[1], Gengyun Li[2], Yupeng Geng[3], Hongxing Yang[1], Yonghong Hu[1] and Ji Yang[2], (1)Shanghai Chenshan Botanical Garden, Shanghai, China, (2)Fudan University, Shanghai, China, (3)Yunnan University, Kunming, China

Introduction: *Alternanthera philoxeroides* is an amphibious invasive weed that can colonize both aquatic and terrestrial habitats. Individuals growing in different habitats exhibit extensive phenotypic variation but little genetic differentiation. The knowledge of transcriptomic reaction underlying phenotypic reaction is critical for understanding the molecular basis of environment-inducible phenotypic changes.

Methods: mRNA and microRNA expression variations in *A. philoxeroides* were characterized throughout the time-courses of pond and upland treatments using RNA-Sequencing. Phenotypic and gene expression responses to contrasting hydrological conditions were compared between *A. philoxeroides* and its alien congener *A. pungens*, which was geographically restricted and terricolous.

Results: 7,805 differentially expressed mRNAs were clustered into 11 transcriptionally coordinated gene groups in *A. philoxeroides*. Functional analysis revealed pathway changes in hormone-mediated signaling, osmotic adjustment, cell wall remodeling and programmed cell death. Meanwhile, 146 differentially expressed microRNAs were found. The microRNA:mRNA pairs potentially associated with plastic internode elongation were identified. The terricolous *A. pungens* displayed limited phenotypic plasticity to different treatments. The interspecific variation in plasticity between *A. philoxeroides* and *A. pungens* was not due to environmentally-mediated changes in hormone levels but to variations in the type and relative abundance of different signal transducers expressed in the target tissue.
Conclusions: This study provides a mechanistic understanding of the biological processes involved in the phenotypic plasticity of *A. philoxeroides*. The genome-wide transcriptional plasticity may alleviate the constraints by genetic impoverishment *in A. philoxeroides* and contribute to its invasion success across diverse habitats.

P1199: Other Plant Species

Mining of the Candidate Genes Involved in Regulating the Chilling Requirement Fulfillment and Bud Dormancy Release of *Paeonia lactiflora* 'Hang Baishao', a Germplasm with the Superb Adaptability during Warm Winters

Jiaping Zhang, Yiping Xia, Danqing Li and Dong Zhang, Zhejiang University, Hangzhou, China

Herbaceous Peony is a world-famous flower and mainly planted in temperate or cool areas. Warm winter in subtropical regions severely hinders "the Southward Plantation of Herbaceous Peony" in the Northern Hemisphere. Studies on the dormancy, chilling requirement (CR) and relevant molecular mechanisms of peony need to be performed. Based on the chilling treatments, the optimal CR of *Paeonia lactiflora* Pall. "Hang Baishao" was 672.00 chilling hours or 856.08 chilling units for achieving the superior sprouting and flowering performances. During the bud dormancy, the ratio of IAA/ABA fluctuated and starch content dropped, while the soluble sugar content and peroxidase activity rose steadily. Transcriptome sequencing were performed for the 'Hang Baishao' buds during the dormancy and sprouting. The differentially expressed genes (DEGs) could be divided into three categories, i.e. DEGs related to environmental response, metabolism, and cell growth. The "difference" in the expression patterns of *SOC1* and *WRKY 33* between two winters, and the "difference" of CR fulfillment periods also between these two winters, represented the interesting congruent relationships. Therefore, they are likely involved in determining the CR fulfillment period of 'Hang Baishao'. The genes related to phytochrome (*PHY*), heat shock protein (*HSP*), osmotin (*OSM*), dehydrin (*DHN*), auxin-repressed protein (*ARP*), repressor of GA (*RGA*), GA20 oxidase (*GA20ox*), peroxidase (*PER*), cyclin (*CYC*) and expansin (*EXP*) expressed actively. This study could contribute to the knowledge of the mechanisms that regulate CR and dormancy characteristics, and may be beneficial for breeding new peony cultivar that have low CR for horticulture use in subtropical regions.

P1200: Other Plant Species

Improvement of a Rubber Tree *Hevea brasiliensis* Genome Sequence

Yasukazu Nakamura[1], **Takako Mochizuki**[1], Hideki Nagasaki[2], Yasuhiro Tanizawa[1], Kohei Ide[3], Teppei Mori[3], Taku Demura[4], Norie Watanabe[3] and Toshihiro Uchiyama[3], (1)National Institute of Genetics, ROIS, Mishima city, Japan, (2)Kazusa DNA Research Institute, Kisarazu city, Japan, (3)Central Research Division, Bridgestone Co., Kodaira city, Japan, (4)Nara Institute of Science and Technology, Ikoma city, Japan

We improved a draft sequence of an estimated 1.4 Gbp genome for *Hevea brasiliensis* PB260 clone, a euphorbiaceous tree producing latex. Bridgestone's research division has been conducting basic research into molecular breeding of *Hevea brasiliensis* to enhance the productivity of natural rubber. To accelerate the research activities, we have decoded a genome for *Hevea brasiliensis*. At the PAG Asia 2016, we presented the optimized scaffolds comprised of 50 k scaffolds. The assembly had N50 size of 120 kb and total scaffold length of 1.6 Gb, containing 227 Mb (14.7%) of unknown nucleotides, in total. In this year, we have improved these scaffolds with the additional PacBio RS II long reads. An improved set of scaffolds comprised of 33 k scaffolds, and has N50 size of 202 kb and total scaffold length of 1.7 Gb, containing 337 Mb (24.1%) of unknown nucleotides, in total. Comparing the previous scaffolds and the current version scaffolds with BUSCO v3, the presence percentage of the gene set of "Embryophyta odb9" is enhanced from 93.8% to 94.1%. We have been continuing the further improvement to decrese the unknown nucleotides. The genome data is expected to facilitate development of *Hevea brasiliensis* breeding technologies that enable us to develop a better breed of the plant that produce high-quality latex with good productivity. The data also accelerate research applications in a variety of fields, including the development of breed with superior disease resistance and environmental stress tolerance.

P1201: Other Plant Species

MarpolBase: Construction of the *Marchantia polymorpha* Genome Database

Yasuhiro Tanizawa[1], Takako Mochizuki[1], Hideki Nagasaki[2], Shohei Yamaoka[3], Ryuichi Nishihama[3], Takehiko Kanazawa[4], Takashi Ueda[4], Katsuyuki T. Yamato[5], Takayuki Kohchi[3] and Yasukazu Nakamura[1], (1)National Institute of Genetics, ROIS, Mishima city, Japan, (2)Kazusa DNA Research Institute, Chiba, Japan, (3)Kyoto University, Graduate School of Biostudies, Kyoto, Japan, (4)Division of Cellular Dynamics, National Institute for Basic Biology, Okazaki, Japan, (5)Kindai University, B.O.S.T., Kinokawa, Japan

The liverwort *Marchantia polymorpha* is a member of a basal land plant lineage and spends most of its life cycle as a haploid gametophyte. Along with its basic characteristics – simple but common with other terrestrial plants, easy observation and manipulation under laboratory conditions make this organism an attractive model for molecular genetics and plant developmental biology.
Previously, we built a wiki-based collaboration site to support genome annotation for *M. polymorpha* by research communities. Here, we report a fully-renewed genome database **MarpolBase** as a data hub for the recently-published draft reference genome (Bowman et al., 2017). To promote the use of the annotated genome data, MarpolBase incorporates a genome browser, sequence similarity searches using BLAST and GMAP, and various utility tools. In addition, we also offer a gene registration system to maintain consistent and systematic nomenclature based on the gene naming guideline (Bowman et al., 2016). MarpolBase will facilitate researches that take advantage of *M. polymorpha*, yet another model organism.

P1202: Other Plant Species

Large Scale Gene Losses underlie the Genome Evolution of Dodder

Guiling Sun[1], Yuxing Xu[2], Hui Liu[2], Ting Sun[1] and Jianqiang Wu[2], (1)Institute of Plant Stress Biology, State Key Laboratory of Cotton Biology, Henan University, Kaifeng, China, (2)Department of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China

Dodders (*Cuscuta* spp., Convolvulaceae) are globally distributed root and leafless parasitic plants that parasitize a wide range of hosts. The physiology, ecology, and evolution of these holoparasites are still poorly understood. A high-quality reference genome (size 264.83 Mb and contig N50 of 3.63 Mb) of *Cuscuta australis* was assembled. Our analysis revealed that *Cuscuta* experienced accelerated evolution, and *Cuscuta* and the convolvulaceous morning glory (*Ipomoea*) shared a common whole-genome triplication event before their divergence. Importantly, *C. australis* genome harbors only 19805 protein-coding genes, and 15.4% of the conserved orthologs in autotrophic plants are lost in *C. australis*, many of which are involved in maintaining autotrophic lifestyle and resistance to stress factors, indicating that gene loss underlies the regressive evolution of *Cuscuta*. The *C. australis* genome provides important resources for studying the evolution of parasitism, regressive evolution, and evo-devo in plant parasites.

P1203: Other Plant Species

***Echinochloa crus-galli* Genome Analysis provides Insight into its Adaptation and Invasiveness as a Weed**

Longjiang Fan, Institute of Crop Science & Institute of Bioinformatics, Zhejiang University, Hangzhou, China

Barnyardgrass (Echinochloa crus-galli) is a pernicious weed in agricultural fields worldwide. The molecular mechanisms underlying its success in the absence of human intervention are presently unknown. Here we report a draft genome sequence of the hexaploid species E. crus-galli, i.e., a 1.27 Gb assembly representing 90.7% of the predicted genome size. An extremely large repertoire of genes encoding cytochrome P450 monooxygenases and glutathione S-transferases associated with detoxification are found. Two gene clusters involved in the biosynthesis of an allelochemical 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) and a phytoalexin momilactone A are found in the E. crus-galli genome, respectively. The allelochemical DIMBOA gene cluster is activated in response to co-cultivation with rice, while the phytoalexin momilactone A gene cluster specifically to infection by pathogenic Pyricularia oryzae. Our results provide a new understanding of the molecular mechanisms underlying the extreme adaptation of the weed.

P1204: Other Plant Species

A Genome Sequence of Clammy Goosefoot, *Dysphania pumilio*

Yongsung Kim and **Jongsun Park**, InfoBoss Co., Ltd., Seoul, South Korea

The genus *Dysphania* R.Br. consisting of 7-10 species endemic to Australia belongs to the tribe *Dysphanieae* in Amaranthaceae. Recent phylogenetic research shows that *Dysphania* is clearly distinguished from *Chenopoium* by having the presence of multicellular glandular hairs and closely related to *Suckleya* and *Cycloloma*. *Dysphania pumilio* (R. Br.) mosyakin & Clemants is the native species to Australia (Scott, 1978; Wilson, 1983; 1984) and it has been spread out to man countries except Africa since 1979. It is strong enough to grow in waste areas or even in tiny space in roadsides. We sequenced genome of *Dysphania pumilio*, invasive species in Korea using HiSeq 4000 with four different libraries. Total length of genome is 365.76 Mb (N50 is 81,724bp), covering 80% of k-mer analysis result (456 Mb). and max length of scaffold is 1.02Mbp. Number of genes predicted by AUGUSTUS is 73,911 among which 43,743 genes (59.18%) have functional domains detected by InterProScan. Functional domain comparison of *Chenopodium quinoa* genome of which genus is neighbor to *Dysphania* and *D. pumilio* presents that 87.24% (5,810) functional domains are shared with both genomes, and 333 functional domains (5.00%) are *D. pumilio* specific. 47,606 Simple sequence repeats (SSRs) were identified from *D. pumilio* genome, presenting that proportions of diSSRs (22.68%) and hexaSSR (4.57%) are higher than those of *C. quinoa* (15.02% and 2.40%, respectively). D. pumilio genome, which is the first genome in *Dysphania* genus will provide detailed landscape of *Dysphania* genomes as well as phylogenetic relationship between *Dysphania* and *Chenopodium*.

P1205: Other Plant Species

Phylogenetic Inferences in New World-Native Chenopodium Allotetraploids from Intron 16A Sequence Data of the Salt Overly Sensitive I Gene

Eric N. Jellen[1], **Miranda McLaughlin**[1], Jacob C. O'Brien[1], Peter J. Maughan[1], Daniel Bertero[2], Ramiro Curti[3], Francisco F. Fuentes[4], Mark A. Tester[5] and David E Jarvis[1], (1)Brigham Young University, Provo, UT, (2)University of Buenos Aires, Buenos Aires, Argentina, (3)University Nacional de Salta, Salta, Argentina, (4)Pontifica Universidad Catolica de Chile, Santiago, Chile, (5)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

The goosefoot genus (*Chenopodium*) is circumpolar in distribution, with taxa native to all of the major continents as well as distant archipelago-endemics. The most famous species are *C. quinoa* or quinoa, the highly nutritious seed crop from the Andean region, and *C. album* or lambsquarters, a notoriously invasive agricultural weed. Quinoa is an allotetraploid ($2n = 36$) composed of an AA subgenome that originated in the Western Hemisphere and a BB subgenome commonly found in Eurasian diploids and hexaploid *C. album*. Quinoa is part of a complex of wild, weedy, and cultivated taxa sharing this AABB genome formula and that are native to North and/or South America. This complex includes weedy North American *C. berlandieri* or pitseed goosefoot; weedy South American *C. hircinum* or avian goosefoot; and Mexican seed (*chia roja*), and vegetable (*huauzontle* and *quelite*) domesticates classified as *C. berlandieri* ssp. *nuttaliae*. Currently, we are using DNA sequence information from intron 16A of the *Salt Overly Sensitive I* (*SOSI*) gene to help clarify relationships among these New World allotetraploids. Our primary objective is to better define the primary gene pool for improving the adaptive characteristics of domesticated quinoa so it can be widely cultivated in lowland environments. Phylogenetic data indicate that the South American taxa fit within a larger clade that includes *C. berlandieri* belonging to the northern Gulf of Mexico coastal ecotype (var. *boscianum*), thus supporting a hypothesis of long-range dispersal to South America of the wild/weedy ancestor that was later domesticated as *C. quinoa*.

P1206: Other Plant Species

***Cercospora beticola* Comparative Genomics sheds new light on the Cercosporin Biosynthesis Pathway**

Melvin Bolton, USDA – Agricultural Research Service, Fargo, ND

Cercosporin is a light-activated secondary metabolite effector produced by many *Cercospora* species that contributes to fungal virulence. The metabolic pathway for cercosporin production has been well-characterized and was previously thought to consist of eight cercosporin toxin biosynthesis (*CTB*) genes. By comparing genome sequences of several ascomycetes, we found that the *CTB* cluster has experienced a number of horizontal transfers across a spectrum of plant pathogenic fungi during evolution. Surprisingly, we noticed that these species also harbored an additional complement of genes on one flank of the established *CTB* cluster. Extensive microsynteny outside of the established cercosporin cluster prompted us to test whether the flanking genes in *C. beticola* are also required for cercosporin biosynthesis. Gene disruption of three genes led to the inability of the fungus to produce cercosporin. Taken together, our findings suggest that the *CTB* cluster includes more genes than previously known. A detailed characterization of these novel genes will be reported.

P1207: Other Plant Species

Whole Genome Differential Expression Signature of Tipburn Resistant Lettuce

Rongkui Han, University of California, Davis, Davis, CA and Richard Michelmore, Genome Center, University of California, Davis, Davis, CA

Tipburn is a worldwide problem in lettuce (*Lactuca sativa*) production. Identified as a localized calcium deficiency, tipburn begins with necrosis at the tip and margins of developing leaves, and results in laticifer rupture and latex coagulation. The visually unpleasant symptoms are unacceptable to consumers and may lead to subsequence microbe-mediated rotting. Many packaging company will reject entire fields of lettuce with more than 5% tipburn incidence. Tipburn phenotype typically manifests a few days before harvest, and shows strong Genotype x Environment interactions, thus necessitating the development of tipburn resistance markers. Using recently developed genomic tools, we were able to obtain high-resolution transcriptomic information from RNA-seq libraries of two crisphead lettuce varieties, El Dorado and Empire, which segregate for tipburn resistance, and four F6 recombinant inbred progenies from a cross between these two varieties. A differential expression analysis identified genes whose expression profiles strongly correlated with tipburn phenotypes and plant genotypes. We clustered co-regulated genes using Weighed Correlation Network Analysis (WGCNA), and pinpointed co-expression clusters highly responsive to tipburn severity. Coupled with ontology and pathway enrichment analyses, the differential expression study not only narrowed down the search scope for tipburn resistance candidate gene, but also helped elucidated possible resistance mechanisms.

P1208: Other Plant Species

Developmental Dynamics of Crassulacean Acid Metabolism (CAM) in *Opuntia ficus-indica*

Nicholas A. Niechayev, Jesse A. Mayer and John C. Cushman, Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV

Large agricultural species of prickly pear cactus (*Opuntia ficus-indica*) show great potential for biofuel, food, and fodder production in semi-arid and arid regions across the globe. However, the developmental basis of *O. ficus-indica* productivity in terms of its transition from C_3 photosynthesis to crassulacean acid metabolism (CAM) remains poorly understood. Utilization of the C_3 photosynthetic pathway might explain the high productivity rates observed for this obligate CAM species. The developemental progression of CAM was assessed in developing pads by titratable acidity, carbon isotopic ratio data, and daily gas exchange measurements. Nocturnal titratable tissue acidity build up began to increase when pads were approximately 5 cm long, and peaked at 20 cm. Isotopic mass spectrometric analysis revealed that the $\delta^{13}C$ ‰ values all averaged between -14.77 ‰ and -15.30 ‰ (typical of CAM plants) regardless of the size category of the pads. Daily 24-hour gas exchange measurements showed that the net daily CO_2 uptake is negative until pads are greater than 10 cm in length, and that primarily all CO_2 uptake occurs at night. Collectively, these results suggest that developing *O. ficus-indica* pads begin as respiring carbon sink tissues that do not perform CAM and then begin to use CAM once pads reach 5 cm in length. C_3 photosynthesis is apparently not contributing to the high productivity rates observed for *O. ficus-indica*. The transcriptome of developing *O. ficus-indica* pads has been sequenced using PacBio IsoSeq cDNA to provide a foundation for understanding the molecular genetic basis of this developmental transition.

P1209: Goats

Genome-Wide Target Enrichment-Aided Chip Design: A 66K SNP Chip for Cashmere Goat

Jinquan LI[1], Xian Qiao[2] and Rui Su[1], (1)Inner Mongolia Agricultural University China, HOHHOT, China, (2)INNER MONGOLIA AGRI UNIVERSITY, HOHHOT, China

Compared with the commercially available single nucleotide polymorphism (SNP) chip based on the Bead Chip technology, the solution hybrid selection (SHS)-based target enrichment SNP chip is not only design-flexible, but also cost-effective for genotype sequencing. In this study, we propose to design an animal SNP chip using the SHS-based target enrichment strategy for the first time. As an update to the international collaboration on goat research, a 66K SNP chip for cashmere goat was created from the whole-genome sequencing data of 73 individuals, consist of two 94K RNA probe libraries. Verification of this 66K SNP chip with the whole-genome sequencing data of 436 cashmere goats showed that the SNP call rates was between 95.3% and 99.8%. The average sequencing depth for target SNPs were 40X. The capture region was shown to be 200bp flanking target SNPs. This chip was further tested in a genome-wide association analysis of cashmere fineness (diameter). Several genetic signals were found marginally associated signaling pathways involved in hair growth. These results demonstrated that the 66K SNP chip is a very efficient and useful tool in the genomic analyses of cashmere goats. The successful chip design shows that the SHS-based target enrichment strategy could be applied to SNP chip design in other species.

P1210: Goats

Hybrid *de novo* Assembly of the Mouflon (*Ovis musimon*) Genome

Ruijun Wang[1], Yanjun Zhang[1], Rui Su[2] and Jinquan LI[2], (1)Inner Mongolia Agricultural University, China, Hohhot, China, (2)Inner Mongolia Agricultural University China, HOHHOT, China

Mouflon (Ovis musimon) with its huge and beautiful horns is considered as one of the ancestors of domesticated sheep. In order to provide novel genome information for sheep breeding and genetic analysis of Bovidae species, we propose to sequence the mouflon genome. We assembled the highly heterozygous mouflon genome using the next-generation sequencing method on an Illumina HiSeq platform. The final

draft genome is approximately 2.87Gb, with contig and scaffold N50 sizes of 110.1 kb and 10.4Mb, respectively. Further analyses predicted 30,160 protein-coding genes in the mouflon genome and 12,111 shared gene families among Bovidae species. The draft mouflon genome assembly will provide data support and theoretical basis for various investigations of the Bovidae species in future.

P1211: Maize, Sorghum, Millet, Sugar Cane, and related

Identification of Combining Ability Patterns for Pearl Millet Hybrid Breeding in West Africa

Felix T. Sattler[1], Anna Pucher[1], Ousmane Sy[2], Ahmad Isaaka[3], Issoufou Kassari Ango[4], C. Tom Hash[5] and Bettina I. G. Haussmann[1], (1)University of Hohenheim, Stuttgart, Germany, (2)Institut Sénégalais de Recherches Agricoles, Thiès, Senegal, (3)Institut National de la Recherche Agronomique du Niger (INRAN), Kollo, Niger, (4)Institut National de la Recherche Agronomique du Niger (INRAN), Maradi, Niger, (5)Previously: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Niamey, Niger

Pearl millet (*Pennisetum glaucum* (L) R. Br.) is a staple cereal in India and Sub-Sahara Africa, and crucial for food security in West Africa (WA). In contrast to India, WA pearl millet hybrid breeding is still in its infancy. Due to its allogamy, pearl millet shows strong admixture in its WA center of origin. Consequently, natural genetically distinct heterotic groups do not exist among WA pearl millets. Developing heterotic groups based on combining ability patterns is therefore required for sustainable hybrid breeding. The objectives of this study were to evaluate heterosis and combining ability patterns among WA pearl millets, and to derive recommendations for future hybrid breeding. Seventeen population varieties from Niger, Nigeria, Mali, Senegal and Mauretania were intercrossed in a diallel mating design. The population hybrids and their parents were tested biennial at nine environments in Niger and Senegal. Average panmictic midparent heterosis (PMPH) for grain yield was 43% (Range: 5-138%). Niger x Senegal inter-country crosses yielded significantly higher than intra-country crosses from those countries (130 $g*m^{-2}$, 125 $g*m^{-2}$; $p<0.05$), including excelling single combinations. Varieties were clearly separated by geographic origin, using a principle coordinate analysis computed from modified Roger's distances (MRD) employing 21 microsatellite markers. However, no clear relationship between MRD and hybrid performance or PMPH was found. Our study indicates great potential for increasing pearl millet productivity in WA using hybrid varieties. Systematically building up heterotic groups based on combining ability patterns towards selected germplasm from Niger *versus* Senegal seems promising, when targeting long-panicle pearl millet hybrids.

P1212: Maize, Sorghum, Millet, Sugar Cane, and related

Identifying Genes that Regulate Panicle Architecture of Sorghum *via* GWAS Coupled with Semi-Automated, High-Throughput Image Analysis

Yan Zhou[1], Srikant Srinivasan[2], Seyed Vahid Mirnezami[1], Aaron M Kusmec[1], Maria G. Salas Fernandez[1], Baskar Ganapathysubramanian[1] and Patrick S. Schnable[1], (1)Iowa State University, Ames, IA, (2)IIT Mandi, Mandi, India

Structural variation in inflorescence traits of cereal crops can influence yield. To identify genetic factors that contribute to the structural variation of inflorescences, a semi-automated phenotyping strategy was applied to extract dimensional features from images of sorghum panicles. Genome-wide association studies (GWAS) detected 31 unique SNPs associated with variation in inflorescence architecture. 13 of these SNPs are located within 350 kb intervals that contain genes whose maize or rice homologs are known to affect inflorescence architecture. 70% of the 31 SNPs associated with panicle architecture are located within genomic regions that exhibit high levels of divergence between converted tropical lines and cultivars, consistent with the hypothesis that these regions were targets for selection during crop improvement.

P1213: Methods: Functional Analysis

A Genome-Wide Sequence-Indexed Collection of Grass Mutants

Amy Cartwright[1], Richard Sibout[2], Debbie Laudencia-Chingcuanco[3], John Vogel[1], Elison Blancaflor[4], Daniel Woods[5], Sebastien Antelme[6], Julian I. Schroeder[7], Michael Raissig[8], Samuel P. Hazen[9], Jonathan J Powell[10], Sarit Weissmann[11], Karen-Beth G. Scholthof[12], Benoit Lefebvre[13], Robert Schmitz[14], Maria Harrison[15], Kirankumar S. Mysore[16] and Prasanna Kankanala[4], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Institut Jean-Pierre Bourgin, Versailles, France, (3)USDA-ARS, Western Regional Research Center, Albany, CA, (4)Noble Research Institute, Ardmore, OK, (5)University of Wisconsin-Madison, Madison, WI, (6)INRA-IJPB, Versailles, France, (7)University of California, San Diego, La Jolla, CA, (8)Stanford University/Dept. of Biology, Stanford, CA, (9)Biology Department, University of Massachusetts, Amherst, MA, (10)The University of Queensland, Brisbane, Australia, (11)Donald Danforth Plant Science Center, Saint Louis, MO, (12)Texas A&M University, College Station, TX, (13)INRA - LIPM, Auzeville, France, (14)University of Georgia, Athens, GA, (15)Boyce Thompson Institute for Plant Research, Ithaca, NY, (16)The Samuel Roberts Noble Foundation, Ardmore, OK

The small grass species *Brachypodium distachyon* (Brachypodium) possesses all the biological attributes of a modern model organism with many traits of interest for the development and improvement of grasses that are of worldwide importance as sources of food, feed, and fuel. In large part due to investments by DOE, numerous *B. distachyon* experimental resources and methods have been developed including an outstanding reference genome sequence, high-efficiency *Agrobacterium*-mediated transformation protocols, efficient crossing methods, a large germplasm collection, 22,000 sequence-indexed T-DNA lines, 54 resequenced accessions and more. In collaboration with researchers in France and the United States, we have created another resource that gives the ability to identify and order plants with mutations in virtually any gene in the genome, which is a powerful research tool that can be used to accelerate research grasses. Through this collaboration, we have created a sodium azide (NaN_3) population, a population of ethyl methanesulfonate (EMS) lines, and a fast neutron radiation (FNR) population. To date, over 800,000 mutations from 859 lines have been identified and are available as a jbrowse track on the Bd21-3 reference genome on Phytozome. There are over 800,000 SNP mutations and over 5,000 small deletions. We currently have over 900 lines in our sequencing queue and have a goal to sequence over 2,000 lines. When completed, this will be an unprecedented resource for forward and reverse genetics. Mutant lines are available on a collaborative basis, email: acartwright@lbl.gov. Brachypodium Resources: https://jgi.doe.gov/our-science/science-programs/plant-genomics/brachypodium/ Bd21-3 reference genome on Phytozome: https://phytozome.jgi.doe.gov/jbrowse/index.html?data=genomes/BdistachyonBd21_3

P1214: Methods: Sequencing

Advanced Genomics Tools for Deep Insights into Complex Genome Systems

Jayson Talag, Arizona Genomics Institute, Tucson, AZ

The Arizona Genomics Institute (AGI) has played significant roles in numerous genome projects over the past 15 years, including Asian and African rice and its 20 wild relatives, maize, wheat, Brachypodium, date palm, sugarcane, citrus, cacao, soybean and its wild relatives, brassicas, tomato, tree nuts, etc. AGI's expertise is not limited to plants, and includes model species like Drosophila (19 genomes), zebra finch, Biomphalaria and nurse shark, dingo, as examples. AGI's philosophy is that the first genome sequence of any species should be as high a quality as possible. To achieve this standard, AGI is currently employing long-read sequencing platform – e.g. PacBio's SEQUEL. Using this instrument, our read lengths Sub read N50 average 20KB, with 23KB on some projects. The average output is ~5Gb/cell, with some over 9Gb/cell. Using this technology to sequence BAC-pool we published two of the highest quality indica rice genomes August of 2016. Theses genome have now been upgraded with the addition of whole genome shotgun PacBio data resulting in near gap free assemblies with less than 20 gaps/genome.

A critical key to our success lies in our ability to isolate high-quality high-molecular weight DNA as initial substrates for library construction. We have found that specific considerations must be addressed to achieve access to genomic substrates (HMW DNA and RNA) for downstream high quality performance. These include defined tissue types and collection protocols, careful extraction procedures and chemistry modifications, advanced purification steps using both chemical and electrophoretic methods, and very stringent quality control measures to assure substrate performance. Our methods have been used to produce high quality substrates for a variety of different applications such as Pacbio, Illumina, RNAseq, 10x Genomics, Dovetail, BAC library construction, etc.

This poster will present our methods and applications to achieve high-quality genome assemblies that will stand the test of time.

P1215: Methods: Transformation

NSF Plant Transformation Workshop

Zhanyuan Zhang, University of Missouri, Columbia, MO

Plant transformation has been a bottleneck in advancing plant functional genomics study and genome editing. Transformation of recalcitrant cereal crop species has been challenging. The National Science Foundation - Plant Genome Research Program sponsors this training workshop. The goal of this workshop is to provide attendees with hands-on experience in Agrobacterium-mediated transformation of cereal crop species. Attendees will have the opportunity to walkthrough advanced protocols for transformation of cereal crops with focus on three recalcitrant cereal species; including Zea mays inbred lines, Sorghum bicolor public genotype, and Brachypodium distachyon public genotype. Trainees will have the opportunity to learn how to utilize plant morphorgenic regulator genes to transform B73 as well as the best practice for cereal transformation. In addition, the workshop will offer two lectures and host a discussion forum. The first lecture will focus on the mechanism of plant somatic embryogenesis whereas the second lecture will center on how to establish and implement cereal transformation systems. The Plant Transformation Core Facility at University of Missouri, Columbia, MO, USA, will host this workshop from July 30 to August 3, 2018. For workshop pre-registration (free), please visit Plant Transformation Core Facility website at **https://plantsciences.missouri.edu/muptcf/workshop** and for any workshop update. Seats are limited. Pre-registration is required by June 20, 2018 to secure your spot and facilitate our workshop organization. Please contact Dr. Zhanyuan J. Zhang, zhangzh@missouri.edu for any workshop related questions.

P1216: Microbes and Pathogens

Sequence Capture NGS for Vertebrate Viruses

Oliver Lung[1], Amber Papineau[1], Michelle Nebroski[1], Todd Wylie[2] and Kristine Wylie[2], (1)Canadian Food Inspection Agency, Winnipeg, MB, Canada, (2)Washington University School of Medicine, St. Louis, MO

The availability of high through-put sequencing technology has improved our ability to detect and characterize pathogens without *a priori* knowledge. However, the abundances of host and environmental nucleic acids can impact the success of accurately identifying low abundant viral nucleic acids. ViroCap is a sequence capture panel which consists of probes designed to enrich sequences representing viral species that infect vertebrates which sequence information is available. ViroCap was designed to capture sequences from viruses representing 190 genera, and has been previously tested on 32 viruses representing 19 genera that affect humans. Here, we further tested ViroCap against a panel of blinded cell culture amplified viruses and clinical/field samples containing another 26 viral species representing 19 genera and 12 families that affects livestock, wildlife and humans. All viral species were accurately identified and a few unexpected viruses were detected. Full or near full genomes were obtained for most tested viral species and enrichment was observed when compared with pre-captured material. These results indicate ViroCap is a useful tool for improving the sensitivity of NGS for identification and sequencing of a broad spectrum of viruses that affect vertebrates.

P1217: Microbes and Pathogens

Multiple Antagonistic Activities against Aflatoxigenic *Aspergillus flavus* by a Biocontrol Yeast *Wickerhamomyces anomalus* WRL-076

Sui-Sheng T. Hua, USDA-ARS, Western Regional Research Center, Albany, CA, **Dan E. Parfitt**, University of California, Davis, CA, Siov Bouy L. Sarreal, USDA-ARS, Western Regional Res. Center, Albany, CA, PerngKuang Chang, USDA, ARS, Southern Regional Research Center, New Orleans, LA, Jiujiang Yu, Agriculture Res. Center, Beltsville, MD and Sui Sheng T. Hua, Siov Bouy L. Sarreal, Dan E. Parfitt*, Perng Kuang Chang, Jiujiang Yu

Aflatoxin (AF) B_1 is very toxic to humans and animals and is the most potent natural carcinogens known. The toxin is often detected in agricultural crops, food products, dried fruits and spices. AF contamination is a serious and recurrent problem and causes substantial economic losses. We observed that when toxigenic strains of *A. flavus* were grown in the presence of an effective biocontrol yeast, aflatoxin production was reduced significantly, fungal cell wall was damaged and spores production was inhibited. The goal of this project was to examine the

molecular mechanism of the antagonistic effects on gene expression of secondary metabolism and morphological development. The global regulatory proteins VeA, a component of the *velvet* nuclear protein complex, is required for the production of aflatoxins and cyclopiazonic acid. The biocontrol yeast repressed veA transcription significantly, which in turn will influence mycotoxin production, and the formation of sclerotia. The results suggest that the yeast, *W. anomalus* has the potential to control aflatoxin and cyclopiazonic acid in the food chain.

P1218: Oilseeds, Sunflower, and related

Improving the Carrot Genome Assembly and Gene Prediction: Strategies to Overcome Challenges from Short Read Genome Assemblies

Hamed Bostan[1], Shelby Ellison[2], Douglas Senalik[2], Phil Simon[2] and Massimo Iorizzo[3], (1)Plants for Human Health Institute, North Carolina State University, Kannapolis, NC, (2)USDA, ARS and University of Wisconsin, Madison, Madison, WI, (3)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC

The release of the carrot genome sequence in 2016 and its gene prediction and annotation have rapidly enhanced molecular and genomic research for this species, changing the nature of research in carrot biology. Research is shifting more towards extensive genetic screening for genome-wide association analysis and functional genomics. However, despite the high-quality of the current carrot genome assembly v2.0 release, improvements are needed due to the multiple challenges associated with short read sequencing data used to develop it. In this work, we aim to i) increase the contiguity of the current carrot genome assembly, ii) anchor a higher proportion of the genome to genetic maps, and iii) improve gene prediction and annotation. Aided by the advancement of new technologies and the availability of long-read sequencing techniques, here we present a preliminary PacBio Sequel genome assembly using FALCON and CANU assemblers in which the contiguity of the assembly (N50) was improved ≈108 fold, compared to the Illumina based published version. In addition, preliminary data to integrate Nanopore and existing paired-end sequences to improve the quality of the genome assembly are presented here. Finally, a comparison between two currently available sets of gene predictions developed with different pipelines indicated that over 6000 gene models have no overlap and only 1,064 genes overlap completely. This motivates the improvement of carrot gene predictions, since a large proportion of gene models are missing or miss-predicted in either one of the sets.

P1219: Other Plant Species

Detection of an *EPSPS* Gene Isoform in Italian Ryegrass Plants Resistant to both Glyphosate and Glufosinate

Parsa Tehranchian[1], Karthik Putta[2], Elizabeth Karn[1], Mithila Jugulam[2], Sarah Morran[1] and Marie Jasieniuk[1], (1)University of California, Davis, CA, (2)Kansas State University, Manhattan, KS

Italian ryegrass (*Lolium perenne* ssp. *multiflorum*) is an economically significant weed in orchards and vineyards of northern California. After two years of ryegrass control with glyphosate and glufosinate at the labeled rates (867 ae ha^{-1} and 984 ai ha^{-1}, respectively), plants in a pear orchard were not effectively controlled by the herbicides. Sanger sequencing of a region of the 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) gene, which encodes the molecular target enzyme of glyphosate, in 10 seedlings, grown from a bulked ryegrass seed sample from the pear orchard, revealed a missense mutation resulting in a Pro_{106}-to-Thr substitution in only one individual. This amino acid substitution is known to cause target-site resistance to glyphosate in ryegrass species. No mutations at position 106 were detected in two resistant individuals, suggesting non-target-site based glyphosate resistance in these plants. Intriguingly, the remaining seven resistant plants exhibited a nucleotide "A" insertion at the second base of codon 106 and the possibility of an *EPSPS* amino acid substitution (Pro_{106}-to-His) conferring resistance to glyphosate. In a second experiment, genomic DNA sequencing of 40 plants from seeds of an individual containing the highest number of seedlings surviving both glyphosate and glufosinate revealed the "A" nucleotide insertion in 35 plants. Five remaining multiple-resistant plants exhibited no mutations at site 106. Sanger sequencing of cDNA and protein expression analysis are underway to further characterize the novel *EPSPS* isoform and the functionality of the resulting protein.

P1220: Rice

Genome-Wide Analyses of Direct Target Genes of Four Rice NAC-Domain Transcription Factors Involved in Drought Tolerance

Pil Joong Chung, Seoul National University, Pyeong Chang, South Korea

Background: Plant stress responses and mechanisms determining tolerance are controlled by diverse sets of genes. Transcription factors (TFs) have been implicated in conferring drought tolerance under drought stress conditions, and the identification of their target genes can elucidate molecular regulatory networks that orchestrate tolerance mechanisms.

Results: We generated transgenic rice plants overexpressing the 4 rice TFs, *OsNAC5, 6, 9,* and *10*, under the control of the root-specific *RCc3* promoter. We showed that they were tolerant to drought stress with reduced loss of grain yield under drought conditions compared with wild type plants. To understand the molecular mechanisms underlying this tolerance, we here performed chromatin immunoprecipitation (ChIP)-Seq and RNA-Seq analyses to identify the direct target genes of the OsNAC proteins using the *RCc3:6MYC-OsNAC* expressing roots. A total of 475 binding loci for the 4 OsNAC proteins were identified by cross-referencing their binding to promoter regions and the expression levels of the corresponding genes. The binding loci were distributed among the promoter regions of 391 target genes that were directly up-regulated by one of the OsNAC proteins in four *RCc3:6MYC-OsNAC* transgenic lines. Based on gene ontology (GO) analysis, the direct target genes were related to transmembrane/transporter activity, vesicle, plant hormones, carbohydrate metabolism, and TFs. The direct targets of each OsNAC range from 4.0-8.7% of the total number of up-regulated genes found in the RNA-Seq data sets. Thus, each OsNAC up-regulates a set of direct target genes that alter root system architecture in the *RCc3:OsNAC* plants to confer drought tolerance. Our results provide a valuable resource for functional dissection of the molecular mechanisms of drought tolerance.

Conclusions: Many of the target genes, including transmembrane/transporter, vesicle related, auxin/hormone related, carbohydrate metabolic processes, and transcription factor genes, that are up-regulated by OsNACs act as the cellular components which would alter the root architectures of *RCc3:OsNAC*s for drought tolerance.

P1221: Sheep

Genome-Wide Association Study Reveals Major Genes Responsible for the Shortened Ear Phenotype in Altay Sheep

Sangang He[1], Wenrong Li[1], Mingjun Liu[1] and Zhihua Jiang[2], (1)Xinjiang Academy of Animal Science, Urumqi, China, (2)Washington State University, Pullman, WA

The shortened ear phenotype appears in many sheep breeds, including Altay, an indigenous Chinese breed of sheep. The objective of this study was to identify major genes underlying this phenotype. A genome-wide association study (GWAS) was performed using 606,006 single nucleotide polymorphisms (SNPs) with DNA samples collected from 26 shortened ear and 29 normal animals. Here we report two major regions: one on ovine chromosome (OAR6) and another on OAR18 that are significantly associated with the shortened ear phenotype (P=3.376e-10 - 9.314e-08). The strong candidate genes include carboxypeptidase Z (CPZ) and H6 family homeobox 1 (HMX1) in the former region and Aryl hydrocarbon receptor nuclear translocator 2 gene (ARNT2) in the latter region, respectively. By sequencing an evolutionarily conserved region (ECR) surrounding HMX1, we found a duplication of 76 bp completely responsible for the shortened ear with a dominant inheritance mode. Our results confirmed the role of the ECR in HMX1 in development of the external ear, which can be used in selection against this phenotype via genetic diagnosis.

P1222: Swine

Transcriptome Responses to Respiratory Virus Infection of Pigs within the Tracheobronchial Lymphnode Following Infection with PRRSV, PCV2 or IAV

Laura C. Miller[1], Damarius S. Fleming[2], Gregory P. Harhay[3], Marcus E. Kehrli[1] and Kelly M. Lager[1], (1)NADC-ARS-USDA, Ames, IA, (2)ORAU, USDA National Animal Disease Center, Texas A&M, Ames, IA, (3)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE

Porcine reproductive and respiratory syndrome virus (PRRSV) is a major respiratory pathogen of swine that has become extremely costly to the swine industry worldwide, often causing losses in production and animal life due to their ease of spread. However, the intracellular changes that occur in pigs following viral respiratory infections are still scantily understood for PRRSV, as well as, other viral respiratory infections. The aim of this study was to acquire a better understanding of PRRS disease by comparing gene expression changes that occur in tracheobronchial lymph nodes of pigs infected with either porcine reproductive and respiratory syndrome virus (PRRSV), porcine circovirus type 2 (PCV2), or swine influenza virus (IAV) infections. The study identified and compared gene expression changes in the TBLN of 16 pigs following infection by PRRSV, PCV2, IAV, or sham inoculation. Total RNA was pooled for each group and time-point (1, 3, 6, and 14 DPI) to make 16 libraries, for analysis by Digital Gene Expression Tag Profiling (DGETP). The data underwent standard filtering to generate a list of sequence tag raw counts that were then analyzed using multidimensional and differential expression statistical tests. The results showed that PRRSV, IAV and PCV-2 viral infections followed a clinical course in the pigs typical of experimental infection of young pigs with these viruses. Gene expression results echoed this course, as well as, uncovered genes related to shared and unique host immune responses to the 3 viruses. By testing and observing the host response to other respiratory viruses, our study has elucidated similarities and differences that can assist in development of vaccine and therapeutics that shorten or prevent a chronic PRRSV infection.