Comparison of Sperm Methylome in Fertile and Infertile Boars Carriers of DNA Translocation

Hervé Acloque

W001: Abiotic Stress

Moving Closer to Genes for Drought Tolerance in Chickpea (*Cicer arietinum*)

Rajeev K Varshney, ICRISAT, Hyderabad, India

Chickpea is the second most important grain legume cultivated by resource poor farmers in the arid and semi-arid regions of the world. Drought is one of the major constraints leading up to 50% production losses in chickpea. Analysis of phenotypic data for 20 drought tolerance traits in 1-7 seasons at 1-5 locations together with genetic mapping data for two mapping populations provided 9 QTL clusters of which one present on CaLG04, referred as "*QTL-hotspot*" explaining 58.20% phenotypic variation for 13 drought tolerance traits. Introgression of the "*QTL-hotspot*", through marker-assisted backcrossing, in different genetic backgrounds has enhanced yield under drought conditations. A combination of two approaches, namely QTL analysis and gene enrichment analysis were used to identify candidate genes in the "*QTL-hotspot*" region. In the first approach, a high-density bin map was developed using 53,223 single nucleotide polymorphisms (SNPs) identified in the recombinant inbred line (RIL) population of ICC 4958 (drought tolerant) and ICC 1882 (drought sensitive) cross. QTL analysis using recombination bins as markers along with the above mentioned phenotyping data splitted the "*QTL-hotspot*" region into two subregions namely "*QTL-hotspot_a*" (15 genes) and "*QTL-hotspot_b*" (11 genes). In the second approach, gene enrichment analysis using significant marker trait associations based on SNPs from the Ca4 pseudomolecule with the above mentioned phenotyping data, and the candidate genes from the refined "*QTL-hotspot*" region showed enrichment for 23 genes. Twelve genes were found common in both approaches. Functional validation using quantitative real-time PCR (qRT-PCR) indicated four promising candidate genes having functional implications on the effect of "*QTL-hotspot*" for drought tolerance in chickpea.

W002: Abiotic Stress

Genetic Mapping and Physiological Breeding Towards Heat and Drought Tolerance in Wheat

Sivakumar Sukumaran, International maize and wheat improvement center, texcoco, Mexico

Wheat represents about 30% of the world cereal area, with 220 million ha cultivated worldwide, often under abiotic stress. Despite advances in understanding the genetic basis of flowering time and disease resistance, the genetic basis of heat and drought tolerance is poorly understood. At present in the CIMMYT physiology group, we have developed and tested several physiological traits related to heat and drought stress that can be used for gene discovery and physiological breeding. Conceptual models for heat and drought stress tolerance to genetically improve wheat through physiological traits were proposed earlier. In this talk, we will focus on the genetic mapping of physiological traits associated with heat and drought tolerance. Heat stress tolerance was measured using spectral indices to detect photo protective molecules, membrane thermostability and stay green traits. Genome-wide association mapping was done on a Wheat Association Mapping Initiative (WAMI) panel for these traits. Also a complementary bi-parental approach was followed on a phenology controlled population—Synthetic × Weebil— to study heat, drought, and heat+ drought tolerance. These populations were genotyped through 90K SNP markers. In addition, advances in genetic gains from physiological trait (PT) crosses based breeding for heat and drought tolerance will be discussed based on the data from international nurseries of CIMMYT.

W003: Abiotic Stress

Prioritizing Root System Architecture QTLs for Marker-Assisted Selection in Durum Wheat

Marco Maccaferri[1], Walid El-Feki[2], Nazemi Ghasemali[3], Silvio Salvi[1], Maria A. Canè[1] and Roberto Tuberosa[1], (1)DipSA - University of Bologna, Bologna, Italy, (2)2Department of Crop Sciences, Faculty of Agriculture, Alexandria University, Alexandria, Egypt, (3)3Department of Agriculture, Haji Abad Branch, Islamic Azad University, Haji Abad, Iran

Optimization of root system architecture (RSA) traits is an important objective for modern wheat breeding. Linkage and association mapping for RSA evaluated in seedlings of two recombinant inbred line populations and one association mapping panel of 183 elite durum wheat (*Triticum turgidum* L. var. *durum* Desf.), respectively, evidenced 20 clusters of QTL for root length and/or number as well as 30 QTLs for root growth angle (RGA). QTLs were mapped on a high-density tetraploid consensus map based on a transcript-associated Illumina 90K SNP assay developed for bread and durum wheat, thus allowing for an accurate cross-referencing of RSA QTLs between tetraploid and hexaploid wheat. Among the main QTL clusters for root length and number highlighted in this study, 15 overlapped with QTLs for multiple RSA traits reported in bread wheat while out of 30 QTLs for RGA, only six showed colocation with previously reported QTLs in wheat. Based on their relative additive effects, allelic distribution in the AM panel and co-location with QTLs for yield and kernel weight, the RSA QTLs have been prioritized in terms of breeding value. Three major QTL clusters for root length and number (*RSA_QTL_cluster_5#*, *RSA_QTL_cluster_6#*, and *RSA_QTL_cluster_12#*) and five QTLs (*QRGA.ubo-2A.3, QRGA.ubo-2B.2/2B.3, QRGA.ubo-4B.4, QRGA.ubo-6A.2* and *QRGA.ubo-7A.2*) for RGA appear particularly valuable for further characterization towards their positional cloning and possible deployment in marker-assisted selection.

W004: Abiotic Stress

The Effects of Elevated Atmospheric CO2 on Root and Shoot Gene Expression and Metabolite Profiles of *Solanum lycopersicum* and *Solanum pennellii*

Sharon Gray, University of California-Davis, Davis, CA, Ted Toal, University of California, Davis, Davis, CA and Siobhan Brady, University of California Davis,, Davis, CA

Elevated CO_2 predicted for the latter half of this century will increase photosynthetic carbon assimilation of C_3 plants, stimulating biomass. Increased root biomass is predicted to improve plant access to water and nutrient resources in the future. However, we do not know the molecular mechanisms and developmental changes that underlie enhanced root biomass in elevated CO_2. We used domesticated tomato (*Solanum lycopersicum*, cv 'M82') and a wild species (*Solanum pennellii*, 'LA0716') as models. We measured the effect of elevated CO_2 on the transcriptome and metabolome of root and shoot tissue, root cellular anatomy and physiological responses over developmental time. Genes that were differentially expressed in response to elevated CO_2 in M82 and *S. pennellii* shoots showed significant enrichment for GO categories related to translation and biosynthesis of ribosome components. Genes that were differentially expressed in response to elevated CO_2 in *S. pennellii* roots were enriched for GO categories including epigenetic regulation of gene expression and chromatin binding. Profiling of primary metabolites also

showed significant variation between CO_2 treatments, species, root and shoot tissue, and dates. Elevated CO_2 significantly increased photosynthesis and root and shoot biomass in both species, and biomass responses tended to be greater in *S. pennellii*. Elevated CO_2 also affected root vascular anatomy, with stronger responses in *S. pennellii* compared to M82. These results shed light on the molecular mechanisms of whole plant responses to climate change, and demonstrate interspecies variation in root responses to elevated CO_2.

W005: Abiotic Stress

Genomic Resources for Trait Discovery and Improved Plant Performances in Soybean

Babu Valliyodan, University of Missouri & National Center for Soybean Biotechnology, Columbia, MO

Soybean is an important cash crop with high protein and oil content and widely used as human food and animal feed. Soybean yield improvement is very much essential due to its market value and the major bottleneck is its narrow genetic basis. Key to the success of breeding programs aiming to crop improvement is the extent of genetic variation present in the germplasm. Advanced sequencing and data analysis techniques will greatly improve our ability to dissect and mine genomes for specific genes underlying key traits and allelic variation. We have selected most diverse cultivated and wild type soybean lines from the U.S. germplasm collection for re-sequencing. Sequencing of 600 genomes were completed from the targeted 1500 lines, and others are in progress. Development of additional reference genomes of selected soybean germplasm is in progress and this information also will be a significant resource for the trait discovery processes. Current knowledge of the genetic diversity, genome structure and evolution, and genome-wide association study of major traits including salinity tolerance for the discovery of novel alleles will be presented. Genotype-phenotype association studies will be highly impacted by the large scale genome resequencing and next generation mapping, and this will help designing genomics-assisted breeding strategies for the development of soybeans with improved plant performances and better yield.

W006: Abiotic Stress

Improving Drought Tolerance in Maize: Transgenic Approaches to Improving Grain Yield Under Water Limiting Conditions

Jeffrey Habben, DuPont Pioneer, Johnston, IA

Lack of sufficient water is a major limiting factor to crop production worldwide, and the development of drought-tolerant germplasm is needed to improve crop productivity. The phytohormone ethylene modulates plant growth and development as well as plant response to abiotic stress. Recent research has shown that modifying ethylene biosynthesis and signaling can enhance plant drought tolerance. First, a transgenic gene-silencing approach was used to modulate the levels of ethylene biosynthesis in maize (*Zea mays*) and determine its effect on grain yield under drought stress in a comprehensive set of field trials. Analysis of yield data indicated that transgenic events had significantly increased grain yield over the null comparators, with the best event having a 9 bushel/acre increase after a flowering period drought stress. Analysis of secondary traits showed that there was a consistent decrease in the anthesis-silking interval and a concomitant increase in kernel number/ear in transgene-positive events versus nulls. Second, we discovered novel negative regulators of ethylene signal transduction in Arabidopsis (*Arabidopsis thaliana*) and maize. These regulators are encoded by the ARGOS gene family. In transgenic maize plants, overexpression of ARGOS genes reduces ethylene sensitivity. Moreover, field testing showed that UBIQUITIN1:ZmARGOS8 maize events had a greater grain yield than nontransgenic controls under both drought stress and well-watered conditions.

W007: Allele Mining

Genomic Selection Meets Transcriptomics: Predicting Quantitative Traits in Cassava

Roberto Lozano, Dunia Pino del Carpio, Marnin Wolfe, Deniz Akdemir and Jean-Luc Jannink, Cornell University, Ithaca, NY

Quantitative traits are affected by many loci, whereas traditional allele-mining approaches operate on few loci at a time. To more effectively mine many alleles and increase their frequency in parallel, we use bioinformatic and statistical approaches borrowed from quantitative genetics. Genomic Selection (GS) methods have been developed to use all markers available across the genome. Previous studies in Cassava showed that even when using a relatively small training population and low-density GBS markers, GS can achieve reasonable prediction accuracies. We are currently interested in using functional annotation to tag SNPs that may be close to genomeregions that have some biological relevance for the trait of interest. As a case study we used data from an experiment evaluating the transcriptome of resistant and susceptible cassava plants after being infected with Cassava Brown Streak Virus (CBSV). Both a resistant and a susceptible cassava genotypes were evaluated at several time points after infection. WGCNA analysis allowed us to partition the set of differentially expressed genes based on their co-expression patterns across the different time points. We used the RNAseq data and in-silico genome wide identification of immunity related genes as input for a multikernel GS model approach. Prediction accuracies were compared between the traditional and the “informed” GS models using cross-validation.

W008: Allele Mining

A New Lab Guide on Genotyping-by-Sequencing for Plant Genetic Diversity Analysis

Yong-Bi Fu, Agriculture and Agri-Food Canada, Saskatoon, SK, Canada

Genotyping-by-sequencing (GBS) has recently emerged as a promising genomic approach for exploring plant genetic diversity on a genome-wide scale. However, many uncertainties and challenges remain in the GBS applications, particularly in non-model species. We develop a genetic diversity focused GBS (gd-GBS) protocol and present it as an easy-to-follow lab guide to assist a researcher through every step of a GBS application from sample preparation, library assembly, sequencing, SNP calling, to diversity analysis. It uses two restriction enzymes to reduce genome complexity, applies Illumina multiplexing indexes for barcoding, and has a custom bioinformatics pipeline (npGeno) for automatic SNP genotyping. In this presentation, I will introduce the new GBS lab guide, illustrate its application, discuss related application issues, and update our current efforts in GBS research. Following these lab bench procedures and using the npGeno pipeline, one could generate genome-wide SNP genotype data for a conventional genetic diversity analysis of a plant species.

W009: Allele Mining

A Next-Generation Marker Genotyping Platform (AmpSeq) in Heterozygous Crops: A Case Study for Marker Assisted Selection in Grapevine

Shanshan Yang[1], Jonathan Fresnedo Ramirez[2], Minghui Wang[3], Linda Cote[4], Peter Schweitzer[4], Paola Barba[5], Elizabeth M. Takacs[1], Matthew D. Clark[6], James J. Luby[6], David Manns[5], Gavin Sacks[4], Anna Katharine Mansfield[5], Jason Londo[7], Anne Fennell[8], David Gadoury[5], Bruce Reisch[1], Lance Cadle-Davidson[7] and Qi Sun[3], (1)School of Integrative Plant Science, Cornell University, Geneva, NY, (2)Institute of Biotechnology, Cornell University, Ithaca, NY, (3)Bioinformatics Facility, Cornell University, Ithaca, NY, (4)Cornell University, Ithaca, NY, (5)Cornell University, Geneva, NY, (6)University of Minnesota, St. Paul, MN, (7)USDA-ARS Grape Genetics Research Unit, Geneva, NY, (8)South Dakota State University, Brookings, SD

Marker assisted selection (MAS) has become widely used in perennial crop breeding programs to accelerate and enhance cultivar development via selection during the juvenile phase and parental selection prior to crossing. Next generation sequencing (NGS) has been widely used for whole genome molecular marker discovery, but it also offers a potential opportunity for tailored molecular marker development with high-throughput and low per-sample cost. This study presents a novel and efficient MAS strategy for molecular marker development incorporating a semi-automated primer design pipeline based on amplicon sequencing (AmpSeq). The results document the development of an MAS package using three relevant traits (flower sex, disease resistance, and modified anthocyanins) in grapevine breeding. Here we discuss several strengths of the amplicon sequencing platform that make this approach of broad interest in diverse crop species: accuracy, flexibility, high throughput, low cost, easily automated analysis, and rapid results.

W010: Allele Mining

3,000 Rice Genomes and SNP-Seek: Rice Allele Mining Made Easy

Ramil P. Mauleon, International Rice Research Institute, Los Baños, Philippines

W011: Allele Mining

Genebanks, Genomes and GWAS

Millicent D. Sanciangco, Nickolai Alexandrov, Grace Lee S. Capilit, Dmytro Chebotarov, N. Ruaraidh Sackville Hamilton, Venice Margarette Juanillas, Hei Leung, Locedie A. Mansueto, Ramil P. Mauleon, Sheila Mae Q. Mercado, Maria Elizabeth B. Naredo, Renato A. Reaño, Victor J. Ulat and Kenneth L. McNally, International Rice Research Institute, Metro Manila, Philippines

One way of addressing the impending global food crisis of sustainably feeding the 10B people in 2050 is by exploiting the genetic diversity in genebanks. To date, only a fraction of rice genetic diversity is being tapped. Recently, the release of the 3,000 rice genomes by IRRI, in collaboration with CAAS and BGI, unlocks new prospects for improving agriculture production and developing rice suitable for the changing climate to support the growing population. To set the foundation for broader and more extensive investigations of discovering novel alleles, we utilized the power of GWAS (genome-wide association study) to detect marker-trait associations using the 3KRG and historical phenotype data, collected over time as genebank characterization data from unreplicated studies for the parental accessions of the diverse genetic stocks in the 3K panel. We detected significant markers associated with several agronomic and yield traits, including culm length, grain length, grain width, 100-grain weight, and days to flowering, concordant with previous studies. Further, we showed the utility of GWAS in unraveling the genetic history of ancestral recombinations from diverse populations. Collaborative research projects at IRRI are underway, examining the responses of rice to biotic and abiotic stresses from the 3K panel using GWAS. Future efforts are directed to validating markers from these studies and identifying donors for population development useful for breeding.

W012: Allele Mining

Transgressive Variation for Yield Components and Dynamic Traits in Jefferson (*Oryza sativa*) x *O. rufipogon* Introgression Lines

Jeremy D. Edwards[1], Georgia C. Eizenga[1], Kathleen M. Yeater[2], Susan McCouch[3] and Anna M. McClung[1], (1)USDA-ARS, Stuttgart, AR, (2)USDA-ARS, College Station, TX, (3)Cornell University, Ithaca, NY

Alleles from wild progenitors of crops can be a source of transgressive variation in modern cultivars. Introgressions from the *Oryza rufipogon* donor (IRGC104591) in an *O. sativa tropical japonica* cultivar (Jefferson) were shown to confer a yield advantage in multi-location field trials. Yield loci were mapped in an advanced backcross population, and subsequently introgression lines (ILs) were developed by backcrossing and molecular selection to capture six promising yield loci. Field studies showed that these ILs have retained the expected yield advantage. To make effective use of the wild alleles and associated transgressive yield increases in breeding programs, it is useful to identify the specific yield components affected in each of the ILs. In the current study, the ILs were evaluated under controlled greenhouse conditions with the objective of determining the effects on yield component traits, the relationships between traits, and the changes in the effects throughout the life cycle. Dynamic traits were assessed by fitting a logistic model to measurements over time. To estimate above ground biomass, a novel image analysis pipeline was developed. Relative to Jefferson, the ILs were slower to develop and produced more biomass. At maturity ILs carrying yield loci were significantly different from Jefferson for: longer flag leaves and panicles (*yld1.1*, *yld2.1*, *yld3.2*, *yld8.1*, *yld9.1*), increased panicle branch number (*yld6.1*, *yld9.1*), and more seed per panicle (*yld2.1*, *yld6.1*, *yld8.1*). These results support a model of transgressive variation where alleles from an agronomically inferior wild donor increase yield by contributing to variation in growth and yield components.

W013: Analysis of Complex Genomes

The Importance of Gene Conversion in Flowering Plants

Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

W014: Analysis of Complex Genomes

Building a Highly Accurate Genomic Diversity Analysis by Comparing Denovo Assembly of Multiple Complex Genomes

Guy Kol, NRGENE Ltd., Ness-Ziona, Israel

Guy Kol, NRGene
Detailed comparison of complex genomes holds great promise of revealing the effect of the exact genome content on the phenotype level. The talk will detail NRGene's Denovo assembly process, provide specific examples from maize and wheat, and demonstrate the value of the "all-to-all" comparison
of multiple Denovo assembled genomes of maize.
Title: Building a highly accurate genomic diversity analysis by comparing Denovo assembly of multiple complex genomes
Submitter's E-mail Address: guy@nrgene.com

W015: Analysis of Complex Genomes
Wild Emmer 10.5 Mb Genome Assembly Reveals Insights on Wheat Domestication
Assaf Distelfeld, Faculty of Life Sciences - Tel Aviv University, Tel Aviv, Israel and The wild emmer wheat sequencing consortium (WEWseq), Tel-Aviv University, Tel-Aviv, Israel
The study of plant domestication attempts to answer fundamental biological and cultural questions and is important for present day plant sciences. Domesticated plants, including wheat, differ from their wild ancestors by a suite of traits called the 'domestication syndrome'. For example, the non-brittle rachis and uniform seed germination are important components of the domestication syndrome in wheat. As a step in understanding crop improvement and domestication, we have recently completed the genome assembly of wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides* accession Zavitan), the direct progenitor of most domesticated wheat varieties. The scaffolds of our *dicoccoides* assembly are >100 times longer than any available wheat assembly (L_{50}=7 Mb N_{50}=414) and were ordered along the 14 chromosomes using genetic linkage data from our recombinant inbred line (RIL) population derived from hybridization of durum wheat (*T. turgidum* ssp. *durum*, cv. Svevo) with Zavitan. Genetic analysis using two years of phenotypic evaluations of domestication related traits revealed 15 loci responsible for the morphological differences between Svevo and Zavitan. Gene diversity in these loci was analyzed using resequencing of a core collection of tetraploid wheat and results will be discussed. The wild emmer wheat assembly is expected to accelerate efforts to identify the genetic control over significant polymorphisms between wild and domesticated wheat.

W016: Analysis of Complex Genomes
Improved Methods for Reassembly and Analysis of the 17 Gb Bread Wheat Genome
Juan D Montenegro[1], Agnieszka Golicz[1], Bhavna Hurgobin[1,2], Huey Tyng Lee[1,2], Chon-Kit Kenneth Chan[2], Paul Visendi[1], Philipp Bayer[1,2], Jacqueline Batley[2] and David Edwards[2], (1)University of Queensland, Brisbane, Australia, (2)University of Western Australia, Perth, Australia
Wheat (*Triticum aestivum*) is one of the most important food crops in the world, and securing an increase in wheat production is essential to feed the growing human population. Building a reliable reference genome will assist in the acceleration of genomics assisted breeding programs, however producing a good genome reference is a challenge in wheat due to polyploidy, size and repeat content of the genome. Here, we have reassembled the wheat genome sequence based on a combination of chromosome arm specific reads, skim based genotyping by sequencing and local synteny. Comparisons with the current wheat reference assembly suggest that we have assembled a larger fraction of the wheat genome with lower redundancy and a greater gene content. Sixteen sequenced wheat cultivars were used to identify sequence diversity within breeding lines which can be applied to accelerate breeding of this important crop.

W017: Analysis of Complex Genomes
Towards Understanding and Harnessing Dosage Variation in Populus
Isabelle M. Henry[1], Matthew S. Zinkgraf[2], Héloïse Bastiaanse[2], Andrew T. Groover[2] and Luca Comai[3], (1)University of California, Davis, CA, (2)US Forest Service, Davis, CA, (3)Plant Biology and Genome Center, UC Davis, Davis, CA
Changes in gene dosage can affect gene function in multiple ways and inducing dosage mutations (insertions and deletions) is a powerful approach to rapidly creating wide phenotypic variation. This approach is rarely used in sexual species because the resulting variation is often meiotically unstable. In clonally propagated species, such as the fast growing tree species Populus, this approach holds many advantages. The dosage variants can be maintained vegetatively, clones can be produced to assess phenotypic consistency and some can be integrated in breeding programs rapidly if their performance warrants it. Using gamma irradiation of pollen grains, we have created a population of ~800 interspecific F1 sibling that vary in their chromosomal composition. Using low-pass whole genome sequencing, we characterized their genomic composition and found that approximately 50% of them carry at least one large-scale insertions or deletions across their genome. Within the population, the entire poplar genome is covered by on average 10 indel mutations, providing sufficient redundancy for effective functional genomic studies. Phenotypic characterization of this population is underway and will provide a powerful platform for both understanding gene function and the effect of gene dosage on phenotypes and on poplar hybrid performance. This resource is publicly available for others to investigate specific traits of interest.

W018: Analysis of Complex Genomes
New Insights on Genome Size Reduction in the High-Polyploid Carnivorous Plant *Utricularia gibba* from a Long-Read, Third Generation Assembly
Victor A. Albert[1], Tianying Lan[1], Alan Cervantes-Perez[2], Rikky W. Purbojati[3], Chunfang Zheng[4], David Sankoff[4], Enrique Ibarra-Laclette[5], Luis Herrrera-Estrella[6] and Stephan Schuster[3], (1)University at Buffalo, Buffalo, NY, (2)LANGEBIO, CINVESTAV, Irapuato, Mexico, (3)Nanyang Technological University, Singapore, Singapore, (4)University of Ottawa, Ottawa, ON, Canada, (5)Instituto de Ecología A.C., Xalapa, Mexico, (6)Laboratorio Nacional de Genomica para la Biodiversidad, Irapuato, Guanajuato, Mexico
Utricularia gibba (the bladderwort, family Lentibulariaceae) is a carnivorous plant with an unusually small nuclear genome, which in our published assembly approximated 82 Mb (Ibarra-Laclette et al., 2013). Despite this small size, syntenic analysis of the *U. gibba* genome revealed 3 lineage-specific whole genome duplications since common ancestry with tomato and grape. Analysis of gene family turnover showed that over

evolutionary time *U. gibba* both gained and lost genes faster than tomato, grape, *Arabidopsis*, or *Mimulus* (Carretero-Paulet et al., 2015). We are further studying these genome evolutionary dynamics through use of PacBio SMRT third generation sequencing. Our PacBio-only de novo assembly (from HGAP.3) has a contig N50 of ~3.4 Mb, with a longest contig of 8.5 Mb. Examination of the contig collection reveals that we have successfully assembled several entire chromosomes, including the 8.5 Mb contig and one as small as 3.8 Mb, nearly 1 Mb shorter than the genome of *E. coli* K-12. These chromosomes are marked on both ends by *Arabidopsis*-type telomere repeats, $(TTTAGGG)_n$. Most notably, the new assembly encompasses a total of ~100 Mb after contaminant filtration, >15 Mb more than our fragmented, published assembly based on 454 and MiSeq reads, which had a scaffold N50 of about 80 Kb. Comparative analysis of the new and old assemblies reveals that most of the difference in genome size relates to repetitive DNA sequences, readily sequenced through by PacBio long reads, that otherwise missed assembly in our short-read approach. While in need of size revision, the *U. gibba* genome remains the smallest reliably assembled angiosperm genome. We will report new findings on centromeric sequences and other repetitive DNA islands, and how these relate to *Utricularia* genome evolutionary dynamics. We will also show how the contiguity of the new assembly permits clear resolution of *U. gibba*'s most recent polyploidy event.

W019: Animal Epigenetics

Comparison of Sperm Methylome in Fertile and Infertile Boars Carriers of DNA Translocation

Herve Acloque, INRA-GenPhySE, Castanet-Tolosan, France

Balanced chromosome translocations is one important cause of reduced fertility for boars and is mostly explained by the production of genetically unbalanced gametes. It has been also hypothesized that balanced chromosomal translocations affect fertility through the induction of epigenetic modifications in germ cells that alter the expression of important genes for meiosis and spermatid differentiation. We previously found abnormal methylation in the GNAS imprinted locus in 2 boars carrying chromosomal translocations. We thus aim to explore the effects of chromosomal translocations on gamete epigenetics by looking at sperm DNA methylation using genome-wide approaches.
By combining Reduced Representation Bisulfite Sequencing and sequencing of immunoprecipitated methylated DNA we compared the sperm methylome of 3 fertile and 3 infertile boars (including 2 with chromosomal translocation). Using these two methodologies we detected the previously described abnormal methylated region on the GNAS locus but we finally identified few DMRs between infertile and fertile boars and none was located close to genes with known function in spermatogenesis.
We then analyzed more in detail the regions located at the vicinity of the translocation breakpoints. We observed a global downregulation of gene expression close to DNA breakpoints and a partial reactivation of genes located on the X chromosome but we do not observed a higher density of DMRs around the translocation breakpoints and none close to the breakpoints. Our results suggest that the Meiosis Silencing of Unsynapsed Chromatin observed in spermatocytes is not associated with local DNA hypermethylation while genes around the translocation breakpoints were downregulated.

W020: Animal Epigenetics

MicroRNAs are Master Regulators of the Bovine Lactation Curve

Eveline M. Ibeagha-Awemu, Dairy and Swine Research and Development Centre, Agriculture and Agri-Food Canada, Sherbrooke, QC, Canada

The production capacity of the bovine mammary gland during a lactation cycle is determined by the number and activities of secretory cells, nutrition, health, and animal's genetics including regulatory factors. The lactation curve (LC) is characterized by rapid increase in milk yield followed by gradual decrease after peak milk (~40-70DIM)until animal is dried off. This study examined the regulatory roles of microRNAs (miRNAs) throughout the LC. Nine high producing Canadian Holstein cows were followed throughout the LC. Milk samples were collected on days 1 (day of calving) (d1), 7, 30, 70, 130, 170, 230, 290 and at dry off when production had dropped to 5kg/day (d5kg) for analyses of milk components and miRNA expression. Milk fat% was constant throughout the LC while protein% increased progressively ($P<0.0000$) to the end of lactation. After Next generation miRNA sequencing, 335 known and 520 novel miRNAs were identified. 23 miRNAs were highly expressed across the LC. Top molecular and cellular functions of top five highly expressed miRNAs (Bta-miR-148A, miR-21-5p, miR-26A, miR-30A-5p and Let-7A-5p) suggest housekeeping roles in cellular growth and development, lipid metabolism, tissue morphology and organ development. Differential gene expression showed that about 400 miRNAs were differentially expressed (DE) ($FDR<0.05$) throughout the LC. The highest number of DE miRNAs (133 to 338) was between d1 and d30/d70/d130/d170/d230/d290/d5kg followed by D7 and d70/d130/d170/d230/d293/d5kg (162 to 205 miRNAs). The least number of DE miRNAs (0 to 87 miRNAs) was between d1 and d7, d30 and d70/d130/d170/d230 and d290 and d5kg. Hierarchical cluster analysis separated the LC into 3 stages: d290/d5kg/d1/d7; d7/d30 and d70/d130/d170/d230. Our data shows that miRNAs play pivotal roles in regulating milk production throughout the bovine LC and in mammary gland growth and productivity.

W021: Animal Epigenetics

Reduced Representation Bisulphite Sequencing of the Cattle Genome Reveals DNA Methylation Patterning

Yang Zhou, Animal Genomics and Improvement Lab, USDA, Beltsville, MD

As a key epigenetic modification, DNA methylation is essential for normal development by regulating processes like gene expression, genomic imprinting, and suppression of repetitive elements. However, DNA methylation in the cattle genome is not well understood. Here we presented the first single-base-resolution maps of DNA methylation in 10 diversely somatic tissues harvested from animals related to L1 Dominette 01449 using reduced representation bisulphite sequencing (RRBS). In total, we observed 1,868,049 high accuracy cytosines that located in the CG enriched regions. Similar to the methylation patterning in other species, the CG context was predominant methylated. Widespread differences were identified between the methylation patterns of CG and non-CG contexts. They were distributed differentially among the genomic features, including the genic regions, CpG island regions, and repetitive sequences. Autocorrelation analysis among adjacent residues illustrated that methylation level of CGs were highly correlated in a region. In contrast, weak or no correlation was found among non-CGs or between CGs and non-CGs. The distinct distribution and low correlation of CG and non-CG methylation suggested that non-CG methylation may be independent from CG methylation and may be mediated by different mechanisms. We also detected 798 tissue-specific differentially methylated cytosines (tDMC) and 131 tissue-specific differentially methylated CpG islands. Combined analysis with the RNA-seq data, we discovered several non-CGs in tDMCs also highly correlated with gene expression, which implied the potential function of non-CG methylation in somatic cells in

addition to pluripotent cells. In summary, we showed that non-CGs exist in bovine tissues and some of them were correlated with tissue-specific functions. This study provided essential information for the cytosine methylation patterning in bovine somatic tissues.

W022: Animal Epigenetics

DNA Methylation and Hydroxymethylation in Early Rabbit Embryos: Consequences of *in vitro* Culture

Mohammed Negash Bedhane, Jigjiga University, Jigjiga, Ethiopia

During the first developmental stages, the genome of the embryo is transcriptionally silent and developmental changes are under the control of maternally inherited factors. Embryonic genome activation (EGA) takes place at later stages (8/16-cell-stage in rabbit) and involves epigenetic modifications. DNA methylation at CpG dinucleotides is an epigenetic mark. CpG methylation is depleted at the early stages and reinstated at the blastocyst stage. Recent findings have shown that demethylation involves the oxidation of methylated DNA into hydroxymethylated DNA. However the role of hydroxymethylation can probably not be restricted to an intermediate in DNA demethylation. Indeed, hydroxymethylation seems involved in gene activation and maintenance of pluripotency, and could therefore be important for EGA. Several studies have suggested that *in vitro* conditions can have a negative impact on epigenetic reprogramming. Therefore, our aim was to investigate the impact of two culture media on methylation and hydroxymethylation in rabbit embryos. To quantify methylated and hydroxymethylated DNA, we implemented an immunofluorescence detection protocol on rabbit embryos cultured in those media until different developmental stages. Our results show that the dynamics of methylation and hydroxymethylation are different between the two culture conditions. Further investigation is needed to compare the *in vitro* cultured embryos to *in vivo* developed ones. To draw solid conclusions, it is advisable to reproduce the experiment with other species such as bovine embryos ahead of further steps to demonstrate on human embryo. Our results will be helpful for the advancement of ART which is challenged by abnormal embryonic development and unsuccessful pregnancy.

W023: Animal Epigenetics

Mapping Parent-of-Origin Effects in the Chicken Transcriptome Using *de-novo* Transcriptome Assembly or Reference Genomes

Fernando Lopes Pinto[1], Mohammad Hossein Banabazi[2], Anna-Maja Molin[1], Elizabeth R. Gilbert[3], Christa Honaker[3], Paul Siegel[3], Göran Andersson[1], Leif Andersson[4] and **Dirk-Jan de Koning**[1], (1)Swedish University of Agricultural Sciences, Uppsala, Sweden, (2)Animal Science Research Institute of IRAN (ASRI), Karaj, Iran, (3)Dept. of Animal and Poultry Sciences, Blacksburg, VA, (4)Uppsala University, Uppsala, Sweden

The Virginia Tech high-growth and low-growth selection lines of White Plymouth Rock chickens have been developed by more than 50 generations of divergent selection for body weight at 56 days of age. These lines have been extensively used for QTL mapping in order to reveal the genetic basis for the huge selection response. To further clarify the mechanisms for how QTLs control phenotypic differences, we measured allelic RNA expression ratios to detect differential expression between alleles derived from the two lines. Using whole transcriptome and genome sequencing, we set out to detect allelic imbalance in the RNA expression of 6 F1 progeny from reciprocal crosses between generation 54 parents from the high and low parental lines. Using RNA samples extracted from liver, hypothalamus and longissimus muscle, we generated circa 250 M (100 bp) RNA sequencing reads per F1 individual as well as a 25-fold coverage DNA sequence of each of the parents. This experimental design allowed us to detect parent-of-origin specific differential expression, since we could predict and test obligatory heterozygous loci in the F1 birds. Using the chicken reference genome sequence and transcriptome we could only map 30% of the RNA reads. We also had an excess (>70%) of allelic imbalance. Using *de-novo* transcriptome assembly we could map > than 70% of the RNA reads and we observed a clear reduction in the allelic imbalance. For about 5400 candidate SNPs we detected statistical evidence for a parent-of-origin effect at $P < 0.001$ for about 500 SNPs. The annotation of the *de-novo* transcriptomes is the next challenge.

W024: Animal Epigenetics

Updates on the French and UC Davis FAANG pilot projects

Huaijun Zhou, University of California, Davis, Davis, CA and Elisabetta Giuffra, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France

W025: Animal Genomics and Adaptation to Climate Change

Workshop Opening Remarks

Susan J. Lamont, Iowa State University, Ames, IA

The framework of this workshop will be introduced in the opening remarks.

W026: Animal Genomics and Adaptation to Climate Change

Genomic Discovery of a Pathway Influencing Adaptation to Temperature Stress in Cattle

Dorian J. Garrick, Department of Animal Science, Iowa State University, Ames, IA, Matt Littlejohn, Livestock Improvement Corporation, Auckland, New Zealand and Steve Davis, LIC, Hamilton, New Zealand

A de novo mutation in Halcyon, a Holstein bull was discovered in New Zealand in daughters of him and his son (Matrix). The mutation resulted in dominant expression of a hairy phenotype in about half the offspring. The hairy offspring sought out water for wallowing and suffered from heat stress in climatic conditions that would normally be considered to be within the thermo-neutral zone. Hairy daughters achieved normal pregnancy rates but at least 1/4 did not lactate and those that did produced lower milk yields than their wild-type sisters. The mutation mapped to PRL known to be involved in mammary gland development and initiation of lactation. A non-synonymous mutation in exon 5 that was novel to affected individuals was found, and predicted to interfere with one of three disuplhide bonds in the prolactin hormone. The cause of the heat stress was not the hairy phenotype, as hairy animals clipped to have slick hair coats demonstrated similarly elevated respiration rates and rectal temperatures as their unclipped hairy sisters. The mutant cows demonstrated much lower sweating rates at 28C than their wild-type sisters. Comparative genomic analyses of yak and bison that are hairy and adapted to cold environments demonstrated unusual variants, not in

PRL but in PRLR. The QTL responsible for the slick-haired phenotype in heat adapted Senepol cattle had also been mapped to the region including the PRLR gene. Sequencing of Senepol PRLR demonstrated a 1 bp deletion. The prolactin pathway is clearly involved in adaptation to temperature-related climate stress.

W027: Animal Genomics and Adaptation to Climate Change

Genome-wide SNP Analysis of Small Ruminant Tolerance to Grazing Stress under Arid Desert Conditions

Ahmed R. Elbeltagy[1], Eui-Soo Kim[2], Joram M. Mwacharo[3], Adel M. Aboul-Naga[4], Barbara Rischkowsky[3], Max F. Rothschild[2], ADAPTmap .[5] and I.S.G.C .[6], (1)Animal Production Research Institute, Ames, IA, (2)Iowa State University, Ames, IA, (3)International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia, (4)Animal Production Research Institute, Cairo, Egypt, (5)Italy, (6)Australia

Desert dwelling animals are exposed to complex biophysical stressors including heat, physical exhaustion, solar radiation, and unavailability of feed and water, which can cumulatively be referred to as grazing stress (GS). GS affect physiological parameters (PP) including rectal temperature, respiration rate, minute ventilation volume, and heat production. Changes in these traits can be used as indicators to assess (in)tolerance to GS. The genomes of desert dwellers, have been exposed to GS for millennia, may carry common chromosomal regions that non-desert animals lack. In this study, we utilized signatures of selection (SS) ("iHS" and "Fst"), and GWAS approaches to analyse genotype data generated using *ovine* and *caprine* 50K Illumina Beadchips to investigate tolerance to GS in desert sheep and goats. For SS analysis, genotype data was from 394 and 366 Egyptian desert sheep and goats, respectively and 895 and 464 non-desert sheep and goats, respectively. For GWAS, PP data from 182 and 151 Barki sheep and goats, were analyzed. Several candidate selection sweep regions were observed in both species. One of the regions on OAR10 (34-43 Mb), spanned genes associated with stress, e.g. tumor suppressors (RB1), angiogenesis and wound healing (FGF). In goats, GRID2 (neurotransmitter receptor affecting neuronal apoptosis), and PDLIM5 (ontogenesis) occurred on CHI6 (26-46 Mb). Multiple candidate QTLs affecting, for instance respiration rate change, within genes playing roles in heat generation (SLC27) and detection of temperature stimulus (NR2F6) were identified on CHI7. Results could prove useful in genomic selection and identification of genes involved in desert stress tolerance.

W028: Animal Genomics and Adaptation to Climate Change

Physiological Responses to Heat Stress in Three Chicken Lines

Ying Wang[1], Perot Saelao[1], Kelly Chanthavixay[1], David A. Bunn[1], Rodrigo Gallardo[2], Susan J. Lamont[3] and Huaijun Zhou[1], (1)University of California, Davis, Davis, CA, (2)University of California Davis, School of Veterinary Medicine, Davis, CA, (3)Iowa State University, Ames, IA

Heat stress is one of the most important environmental factors negatively impacting poultry production and health. Genetics are an important contributor in responding to heat stress. Three different chicken lines, two genetically distinct highly inbred lines (Leghorn and Fayoumi) and one commercial egg-laying line (Hy-Line Brown), were characterized for phenotypic differences in response to heat stress. At 14d of age, birds were treated at 35C with 60% relative humidity or at 22C as a control. For the heat-treated group, birds were inoculated at 21d with 10^7 EID50 Newcastle disease virus (NDV) La Sota strain to investigate the effects of both heat stress and NDV infection. Body temperature and blood components were measured using the iSTAT at four stages: Pre-heat, Acute Heat, Chronic heat 1 and 2 (AH, CH1, CH2, at 4h, 7d and 10d post heat treatment, respectively). Body temperature differences between treated and non-treated group were significant at CH2 for both inbred lines and at CH1 in Hy-Line. Most iSTAT parameters were significantly changed with heat treatment in Fayoumis and Hy-Line birds, but not in Leghorns. The heat-resilient Fayoumis had low sodium and glucose levels, high pH and low Hct and Hb. Physiological measurements of Hy-Line birds were more diverse and altering directions or magnitude of these measurements were mostly opposite to Fayoumi birds. This study provides a foundation for characterizing genetic resistance to heat stress in chickens and the biomarkers of physiological responses could serve as potential candidates for predicting heat tolerance.

W029: Animal Genomics and Adaptation to Climate Change

Quantitative Trait Loci Identified for Body Temperature, Body Weight, Breast Yield, and Feed Digestibility in an Advanced Intercross Line of Chickens Under Heat Stress

Angelica Van Goor[1], Kevin J. Bolek[2], Mike E. Persia[3], Chris Ashwell[4], Max F. Rothschild[1], Carl J. Schmidt[5] and Susan J. Lamont[1], (1)Iowa State University, Ames, IA, (2)University of California, Davis, Davis, CA, (3)Virginia Polytechnic Institute and State University, Blacksburg, VA, (4)North Carolina State University, Raleigh, NC, (5)University of Delaware, Newark, DE

Losses in poultry production due to heat stress have negative economic consequences. It has been previously demonstrated that host genetics partially determines response to heat in poultry. The F18 and F19 generations of a broiler (heat susceptible) by Fayoumi (heat resistant) advanced intercross line (AIL) were used to facilitate fine mapping of quantitative trait loci (QTL). Birds were exposed to daily heat cycles from 22 to 28 days of age, and phenotypes were measured in three major phases: pre-heat, first day of heat treatment, and one week of heat treatment. Cloacal temperatures were measured at all three phases; body weight at pre-heat and one-week heat phases. Breast muscle yield was calculated as a percentage of body weight, and Ileal feed digestibility was assayed from digesta collected from the ileum both at day 28. Birds were genotyped using the 600K Affymetrix chicken array and a genome wide association study was completed using Bayesian analyses and 1-Mb windows. Heritability estimates were low to moderate, ranging from 0.03-0.35. We identified QTL for: body temperature on GGA14, 15, 26, and 27; body weight on GGA1-8, 10, 14, and 21; feed digestibility on GGA19-21; and for breast muscle yield on GGA1, 15, and 22. A very large QTL for breast yield on GGA1 explained more than 24% of the genetic variation. Significant QTL were found for each of the measured traits, indicating the feasibility of using genomic selection to improve performance of animals raised under hot conditions. Support: USDA-NIFA-AFRI grant and Hatch project #5358.

W030: Animal Genomics and Adaptation to Climate Change

Transcriptomic Analysis of Broiler Chickens Under Normal and Heat Stress Conditions

Elizabeth M. Pritchett[1], Susan J. Lamont[2], Michael E. Persia[3], Chris Ashwell[4], Max F. Rothschild[2] and Carl J. Schmidt[1], (1)University of Delaware, Newark, DE, (2)Iowa State University, Ames, IA, (3)Virginia Polytechnic and State University, Blacksburg, VA, (4)North Carolina State University, Raleigh, NC
As global climate changes, it is important to improve the resilience of farm animals to the anticipated increases in average temperature and frequency of heat waves. One approach is to identify alleles that improve animals' ability to survive in the face of heat stress. Our laboratories are using transcriptomics to identify genes in the chicken that are sensitive to heat, with the underlying hypothesis that such genes play a role in controlling the response of the animal to this stress. Current studies are defining transcriptome responses of skeletal muscle, hypothalamus and pituitary in the Ross 708 broiler line to cyclical daily heat stress. The presentation will focus on implications of these analyses from both control and heat stress conditions.

W031: Animal Genomics and Adaptation to Climate Change

Genomic Footprints of Natural Selection for Ecological Adaptation in Ethiopian Cattle Populations

Kwan-Suk Kim, Chungbuk National University, Cheongju, South Korea, Zewdu Edea Bedada, Chungbuk National University, Chungcheongbuk-do, South Korea and Tadelle Dessie, International Livestock Research Institute, Addis Ababa, Ethiopia
Natural selection has shaped the genetic diversity of African indigenous cattle genetic resources to adapt to their local environments and to changing environmental conditions. This diversity is detectable in the form of diseases resistance and tolerance to climatic conditions, but has not yet been well understood at the molecular level. Genomic scans for ecological adaptation between breeds adapted to specific ecological conditions could help to elucidate the genomic basis of local adaptation and to understand the genomic changes associated with the responses of organisms to their a biotic environments. We have sampled indigenous Ethiopian cattle populations from arid/semi-arid and highland agro-ecologies and identified highly differentiated loci using Bovine SNP 80 K chip derived from the *indicine* breeds. Some of the highly differentiated loci were positioned within biologically important genes for skin pigmentation, hypoxia, feeding behavior, heat stress responses, immunity, body weight and meat quality. The follow-up study will include Ethiopian indigenous sheep populations collected in the same ecological zones/geographical regions. Positive selection of animals under similar local conditions might have accumulated key genomic components overlapping between species.

W032: Animal Genomics and Adaptation to Climate Change

Workshop Closing Remarks

Susan J. Lamont, Iowa State University, Ames, IA
Highlights of this workshop will be summarized in the closing remarks.

W033: Aquaculture

Editing Fish Genome With CRISPR

Wenbiao Chen, Vanderbilt University School of Medicine, Nashville, TN
The advent of CRISPR genome editing has revolutionized genetic analysis. We have adapted CRISPR for targeted gene disruption in zebrafish. We show that delivery of in vivo synthesized single guide RNA (sgRNA) and mRNA for a codon optimized Cas9 by microinjection results in efficient biallelic gene inactivation in somatic cells, allowing rapid functional assessment in the injected fish. In addition, several genes can be disrupted simultaneously by injecting multiples sgRNAs. Furthermore, we show that transgenic expression of sgRNA and Cas9 is sufficient to cause biallelic mutagenesis, and tissue specific and/or inducible expression of Cas9 affords conditional mutagenesis. Finally, I will discuss our work on generating point mutations and targeted integrations using CRISPR.

W034: Aquaculture

Functional Studies in Atlantic Salmon (Salmo salar L.) Reveals Candidates for Sterility Vaccines

Anna Troedsson-Wargelius, Institute of Marine Research, Bergen, Norway
Recent biotechnological innovations such as the CRISPR-Cas9 methodology, have allowed applying genetic engineering approaches also to non-model organisms, including studies on gene function in Atlantic salmon. We have explored this methodology with the aim to target one of the major sustainability problems in salmon aquaculture: sexually mature escapees. Genetic introgression into wild populations is currently one of the factors limiting the expansion of the Norwegian salmon industry. To address this problem, we are investigating the possibility to induce sterility by vaccination against factors mediating germ cell survival in salmon. We have used CRISPR-Cas9 technology to elucidate the function of candidate genes in germ cell survival. Due to the long generation time of salmon we chose to analyze complete loss of function in F0. To avoid analysis of mosaic individuals, we simultaneously induced CRISPR-Cas9-mediated mutations in the *albino* (*alb*) and in the target gene. We observed that complete loss of pigmentation indicated bi-allelic disruption of *alb* but also of the second gene that was targeted. This methodology allowed producing germ cell-free salmon in F0. Different from medaka and zebrafish but similar to the loach, male and female somatic sex differentiation of the somatic gonadal elements did take place in germ cell-free salmon. We are also now following growth in germ cell free fish to evaluate if this type of sterility is suitable for the aquaculture industry. In addition we are at this time investigating a sterility vaccine using this target.

W035: Aquaculture

Regulatory Approval of Genetically Engineered AquAdvantage Salmon

John Buchanan, Center for Aquaculture Technologies, San Diego, CA
More than 30 years after the first reports of successful gene transfer in animals, regulatory approval for the commercial application of the first transgenic food animal was recently realized. AquaBounty Technologies Inc. was successful in securing Unites States Food and Drug Administration (US FDA) approval for the cultivation and sale of a line of transgenic Atlantic salmon, with the brand name AquAdvantage Salmon. This salmon line was engineered to contain a single additional copy of a salmon growth hormone gene isolated from the Chinook salmon genome, under the control of regulatory elements from an ocean pout anti-freeze protein gene. The phenotype associated with this

transgenic line was remarkable, with transgenic fish exhibiting a growth rate to market size more than double that of non-transgenic siblings. In addition, the transgenic fish exhibited a significant improvement in feed efficiency. The commercial approval of this line of salmon indicates that the US FDA framework for approval of transgenic food animals is navigable. This opens exciting possibilities for future commercial deployment of genetically engineered food animals to increase the productivity and sustainability of animal agriculture.

W036: Aquaculture

Genomic Selection For Bacterial Cold Water Disease Resistance Reveals Large Within-Family Variation That Cannot Be Exploited In Traditional Family-based Selective Breeding In Rainbow Trout

Roger L. Vallejo, USDA-ARS-NCCCWA, Kearneysville, WV

Selective breeding is an effective strategy to improve resistance to specific pathogens, and thus has the potential to mitigate antibiotic use in aquaculture. Large family sizes of aquaculture species permits family-based selective breeding programs, but the need for specific-pathogen-free nucleus populations precludes the ability to exploit within-family genetic variation. Genomic selection (GS) simultaneously incorporates dense SNP marker genotypes with phenotypic data from related animals to predict animal-specific genomic breeding value (GEBV), which circumvents the need to measure the disease phenotype in potential breeders. Here, using a commercial rainbow trout (*Oncorhynchus mykiss*) population, we provide empirical data demonstrating the power of GS to exploit within-family genetic variation for bacterial cold water disease (BCWD) resistance. Animal-specific GEBV was derived for pathogen-naïve breeding candidates based solely on genotypic data after estimating SNP effects in a related training population. Pairs of full-sib sisters with divergent GEBVs for BCWD resistance were mated to a common "random" sire to test the hypothesis that progeny from high-GEBV dams exhibit greater BCWD resistance than progeny from low-GEBV dams. Across 46 paternal half-sib families, the average survival following experimental BCWD challenge in progeny from the high-GEBV dams (44.5%) was 79% greater than in progeny from the low-GEBV dams (24.9%). This response to one generation of GS is attributable solely to within-family variation that cannot be exploited in traditional family-based breeding programs. Furthermore, genome-wide association analysis in this population identified moderate- to large-effect QTL distributed among 17 chromosomes explaining up to 53% of BCWD genetic variance. Similar genetic architecture for BCWD was observed in other unrelated rainbow trout population.

W037: Aquaculture

Weighted ssGBLUP Improves Genomic Selection Accuracy for Survival in a Rainbow Trout Population

Breno O. Fragomeni, University of Georgia, Athens, GA

The objective of this study was to compare methods for genomic evaluation in a Rainbow Trout (Oncorhynchus mykiss) population for survival when challenged by*Flavobacterium psychrophilum*, the causative agent of bacterial cold water disease(BCWD). The used methods were: 1)regular ssGBLUP that assumes all SNPs have the same variance; 2)weighted ssGBLUP(wssGBLUP) that gives more weight to SNPs that explain considerable portion of the genetic variance; both GEBV and SNP effects are updated iteratively; 3)BayesB that performs variable selection. The benchmark method was traditional BLUP. While BayesB used only phenotypes of genotyped animals, ssGBLUP considered all phenotypes, genotypes, and pedigrees jointly. Phenotypes for survival days and pedigrees were available for 4,004 and 5,104 individuals, respectively; whereas 2,490 animals were genotyped for 41k SNPs. Accuracy was the correlation between adjusted phenotype and GEBV divided by the square root of heritability, in a 5-fold cross validation. Accuracies obtained from BLUP, ssGBLUP, BayesB, and wssGBLUP were 0.19, 0.45, 0.61 and 0.63, respectively. The increased accuracy obtained by weighting SNP differently can be explained by presence of large QTL, by population structure (which can lead to fewer independent chromosome segments that may have larger effects), or by the small size of the genotyped population. In Manhattan plots, iteration number 3 had some SNPs explaining more than 7% of the genetic variance of the trait, meanwhile regular ssGBLUP had all regions explaining less than 0.25%. WssGBLUP utilizes all the available information, is simple to apply for complex models, and was the most accurate method in this study.

W038: Aquaculture

Genome Scan for Selection Signatures in Atlantic Salmon Populations Using a High Density SNP Array

María Eugenia López Dinamarca, University of Chile, Santiago, Chile

Detecting selection in domesticated species can provide insight into the biological process of artificial selection and the causal genes underlying phenotypic variation. In some cases artificial selection can drive genetic changes on a short time scale in domesticated animals. Atlantic salmon (*Salmo salar*) represent an ideal model for studying the genomic response to selection because some farmed populations of extant wild stocks have experienced intense artificial selection for traits such as growth for over 40 years.

We used a high density SNP chip to screen 151,509 SNPs for selection signatures in two independently domesticated strains and their respective wild ancestor populations. The genome-wide genetic differentiation between the domestic and wild populations as measured by F_{ST} was 0.16 for Canadian populations and 0.08 for Scottish populations. We performed F_{ST} outlier and cross-population extended haplotype homozyogisty tests to identity signals of artificial selection in the genome. The F_{ST} scans produced a subset of 160 and 93 outlier SNPs showing evidence of selection for Canadian and Scottish stocks respectively.

These identified outlier loci illustrate the usefulness of a high density, genotyping chip and provide a basis for further studies aimed at detecting genes involved in biological processes relevant to domestication and selection in Atlantic salmon.

W039: Aquaculture

The vgll3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (Salmo salar L.) Males

Fernando Ayllon, Institute of Marine Research, Bergen, Norway, Norway

Wild and domesticated Atlantic salmon males display large variation for sea age at sexual maturation, which varies between 1-5 years. Previous studies have uncovered a genetic predisposition for variation of age at maturity with moderate heritability, thus suggesting a polygenic or complex nature of this trait. The aim of this study was to identify associated genetic loci, genes and ultimately specific sequence variants conferring sea age at maturity in salmon. We performed a genome wide association study (GWAS) using a pool sequencing approach (20 individuals per river and phenotype) of male salmon returning to rivers as sexually mature either after one sea winter (2009) or three sea winters

(2011) in six rivers in Norway. The study revealed one major selective sweep, which covered 76 significant SNPs in which 74 were found in a 370 kb region of chromosome 25. Genotyping other smolt year classes of wild and domesticated salmon confirmed this finding. Genotyping domesticated fish narrowed the haplotype region to four SNPs covering 2386 bp, containing the vgll3 gene, including two missense mutations explaining 33-36% phenotypic variation. A single locus was found to have a highly significant role in governing sea age at maturation in this species. The SNPs identified may be both used as markers to guide breeding for late maturity in salmon aquaculture and in monitoring programs of wild salmon. Interestingly, a SNP in proximity of the VGLL3 gene in humans, has previously been linked to age at puberty suggesting a conserved mechanism for timing of puberty in vertebrates.

W040: Aquaculture

Genome-Wide Association Study for Identifying Genome Loci That Affect Fillet Yield in Rainbow Trout (Oncorhynchus mykiss)

Dianelys Gonzalez-Pena, USDA-ARS-NCCCWA, KEARNEYSVILLE, WV

Fillet yield (FY, %) is an economically important trait in rainbow trout aquaculture that reflects production efficiency. Despite that, FY has not received much attention in breeding programs because it is costly to measure and difficult to select on, limiting the genetic progress in traditional selection programs. The recent development of high-density SNP array for rainbow trout has provided the needed tool for studying the underlying genetic architecture of this trait. Here, we conducted a genome-wide association study (GWAS) for FY using the synthetic population of rainbow trout developed at the National Center for Cool and Cold Water Aquaculture. The GWAS analysis was performed using the weighted single-step GBLUP method. Phenotypic records of 1,487 fish from ~ 300 full-sib families, in three successive generations, sacrificed at ~1.5 kg were analyzed with genotype data from 1,183 fish (875 with FY phenotypes and 308 non-phenotyped ancestors). A total of 38,107 effective SNPs were analyzed in a univariate model with hatch year and harvest group as fixed effects and animal as a random effect. Two non-overlapping windows of 25 SNPs located on chromosome Omy28 were responsible for 1.4% and 1.1% of the genetic variability for FY. Among the detected SNPs, 62% were in genes that participate in insulin binding, lipid metabolism homeostasis, cations channel activity, and ATP metabolism and 38% were near these genes. Further studies will evaluate the effects of these markers on growth performance and fillet quality in rainbow trout.

W041: Aquaculture

A Genome-Wide Association Study for Low Oxygen Tolerance in Catfish using the 250K SNP Array

Xiaozhu Wang, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

Low-oxygen tolerance is a major performance trait for aquaculture because hypoxia can cause major mortalities. In addition, use of aerators increases production costs. Catfish, a dominant aquaculture species in the US, is highly tolerant to low oxygen, but enormous losses are still caused by exposures to hypoxia each year. Therefore, improving low-oxygen tolerance through genetics is of great interest to the catfish industry. To identify loci associated with hypoxia tolerance, in this work we conducted a genome-wide association study (GWAS) using the catfish 250K SNP array with 376 channel catfish from six different strains, including Marion Random, Marion Select, Thompson, Kansas Random, Kmix (Kansas × Kansas Select), and 103KS (NWAC 103 × Kansas Select). A genomic region on linkage group 6 was found to be significantly associated with low-oxygen tolerance. In addition, six additional suggestively associated QTL regions were identified on linkage group 4, linkage group 5, linkage group 10 and linkage group 12, respectively. The hypoxia associated SNPs and candidate genes discovered in this study provided new insights into selection strategies for improving catfish brood stocks in aquaculture. Furthermore, some of these candidate genes are involved in MAPK, PI3K/AKT and mTOR signaling pathways, which are essential to hypoxia-mediated apoptosis, cell proliferation and tumor angiogenesis in cancer tissues, suggesting evolutionary conservation of functions of these genes across a broad range of species.

W042: Aquaculture

Candidate genes for ESC Disease Resistance of Catfish as Revealed by a Genome Wide Association Study

Tao Zhou, Auburn University, Auburn, AL

Enteric septicemia of catfish (ESC), caused by the bacterial pathogen *Edwardsiella ictaluri*, is the leading disease problem for the catfish industry. It causes tens of millions of dollars of economic losses each year. Channel catfish (*Ictalurus punctatus*) is generally susceptible to the disease while blue catfish (*Ictalurus furcatus*) is generally resistant. Therefore, the hybrid catfish is a great model system to study ESC resistance because of the strong phenotypic contrast between channel catfish and blue catfish. In this study, 499 individuals of the fourth generation of backcross progenies of hybrid catfish were used to conduct genome wide association study (GWAS) for ESC resistance. A genomic region on linkage group 1 was found significantly associated with ESC disease resistance. Within this region, 14 genes have known functions in immunity, including *wwp1*,*inhba*, *fzd8*, *klf6*, *nck1*, *agtr1*, *racgap1*, *nlrp12*, *trpc1*, *acbd5a*, *apbb1ip*, *myo3a*, *stat2*, *nlrc3*. Among the 14 genes, 8 genes or their paralogs were found differently expressed between ESC resistance and sensitive catfish by RNA-seq analysis of previous study. The genes within the significant QTL on LG1 were known to be related to immune functions, such as phagocytosis, defense response, stimulus response, T cell differentiation, B cell differentiation, macrophage development, antigen processing and presentation. In addition to the significant QTL, two suggestively associated QTL regions were identified on linkage group 12 and 16. It appears that the identified QTLs are limited to specific families of F4 fish, consistent with the segregation of the genomic regions derived from the original blue catfish parent.

W043: Aquaculture

The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates

Yann Guiguen, INRA-SCRIBE, Rennes, France

Vertebrate evolution has been shaped by several rounds of whole-genome duplications (WGDs) that are often suggested to be associated with adaptive radiations and evolutionary innovations. Due to an additional round of WGD, the rainbow trout genome offers a unique opportunity to investigate the early evolutionary fate of a duplicated vertebrate genome. We show that after 100 million years of evolution the two ancestral subgenomes have remained extremely collinear, despite the loss of half of the duplicated protein-coding genes, mostly through pseudogenization. In striking contrast is the fate of miRNA genes that have almost all been retained as duplicated copies. The slow and stepwise rediploidization

process characterized here challenges the current hypothesis that WGD is followed by massive and rapid genomic reorganizations and gene deletions.

W044: Aquaculture

A New and Improved Rainbow Trout (Oncorhynchus mykiss) Reference Genome Assembly

Guangtu Gao, USDA-ARS-NCCCWA, Kearneysville, WV

In an effort to improve the rainbow trout reference genome assembly, we re-sequenced the doubled-haploid Swanson line using the longest available reads from the Illumina technology; generating over 510 million paired-end shotgun reads (2x260nt), and 1 billion mate-pair reads (2x160nt) from four sequencing libraries with fragment sizes ranging from 800bp to 15Kb. We then generated a 2.17Gb genome assembly, containing 139,726 scaffolds with N50 greater than 1.7Mb using DeNovoMAGIC, the advanced genome assembly pipeline from NRGene. To generate a dense genetic map for each chromosome, we genotyped more than 4,000 pedigreed fish with the Affymetrix 57K SNP array and with RAD SNPs. Using these genetic maps' information we anchored and ordered ≈80% of the assembly scaffolds within chromosomes. Our initial analysis indicates that the new physical genome sequence assembly and its integration with the new genetic maps results in major improvements over the current rainbow trout reference genome in all categories; including genome coverage, scaffold size, precision, and anchoring and ordering of the scaffolds onto chromosome maps. The assembly can be further improved by merging of neighboring scaffolds using overlapping long sequences from other trout assemblies or long reads sequencing technologies.

W045: Aquaculture

Progress of the Shrimp Genomic Sequencing Project

Jianhai Xiang, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China

Penaeid shrimp are most economically important marine aquaculture species in China and also in the world. Genetic information, especially whole genome sequence is necessary to better understand the biological essence, and to promote the domestication and genetically improvement in shrimp. However, because its genome is large and complex in structure, estimated approximately 80% repetitive sequences, it is very difficult to sequence and good assemble. Both genomes of the Pacific white shrimp*Litopenaeus vannamei* and Chinese shrimp *Fenneropenaeus chinensis* have been sequenced and assembled for many years funded by grants of the National High Technology Research and Development Program of China. Using Next Generation Sequencing (NGS) technologies, different insert-size pair-end and mate-pair Illumina shotgun libraries were constructed, a total of 828GB and 530GB data were generated, covered 318X and 280X genome size of the two important cultured species of shrimp respectively. PacBio single molecule sequencing was adopted also and more than 70 GB data were generated for assisting assembling in each genome. A set of assembly methods were developed to solve the problems brought by high repetition and heterozygous character of shrimp genome, two preliminary genome drafts were constructed. A set of strategies have been utilized to facilitate assembly. The *L. vannamei* genome scaffold N50 was up to 123.98Kb; and the assembled *F. chinensis* genome scaffold N50 reached 154.2Kb. Paired BAC-end sequencing was conducted, 28,000 BAC ends were obtained. A high-density genetic map of *L. vannamei* has been developed, on which 6,359 SNP markers were mapped to the 44 linkage groups spanning 4,243 cM. Two dozen of transcriptomes were sequenced and all reads were assembled into 117,539 and 188,201 unigenes in *L. vannamei*and *F. chinensis*, at least 95% unigenes were mapped into the assembled genomes. Annotation of the genome of both shrimp is going on. These genome sequences and resource will facilitate the understanding of shrimp and crustacean genome, and should improve the genetic breeding in shrimp.

W046: Aquaculture

Comparative Transcriptome Analysis of the Swimbladder Reveals Expression Signatures in Response to Hypoxia in Channel Catfish, Ictalurus punctatus

Qiang Fu, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

Low oxygen (hypoxia) can have adverse impacts on aquaculture fish production. Catfish is the leading aquaculture species in the US. In spite of the relatively high levels of tolerance of catfish against low oxygen, hypoxia can still lead to huge economic losses due to heavy mortalities. Studies on low oxygen tolerance, therefore, are important for aquaculture. Fish swimbladder has long been believed to be the homolog of the tetrapod lung, but the molecular evidence is still lacking. In this study, we conducted RNA-Seq analysis using swimbladder samples of catfish under low oxygen and normal conditions to determine if swimbladder was involved in respiration, and to reveal genes, their expression patterns and pathways involved in response to hypoxia in catfish. A total of 529 million 100bp paired-end reads were generated. The short reads were assembled by Tophat and Cufflinks software into 170,838 contigs using the reference genome sequence as a guide. The average contig length was 3,215 bp and contig N50 was 6,139 bp. The differentially expression genes (DEGs) were identified with respect to oxygen levels and body size. Subsequent function analysis revealed that many DEGs under hypoxia were involved in carbon dioxide transportation, respiration metabolism, stress response, collagen, and immune responses. In addition to these genes, DEGs between fingerling and adult also included adhere junction, growth regulation and signal transduction. Taken together, these results suggested that the teleost swimbladder indeed have functions for respiration, and that the teleost swimbladder resembles in function to the tetrapod lung.

W047: Aquaculture

The Catfish MicroRNAome: Identification, Annotation and Expression Profiling in Response to Bacterial Infections and Hypoxia Stress

Shikai Liu, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

MicroRNAs (miRNAs) are a class of small noncoding RNAs that play important regulatory roles in multicellular organisms. miRNAs interact with mRNAs by complementing with their 3'-UTRs to suppress translation or initiate mRNA degradation. While a large number of RNA-Seq have been conducted in catfish from various tissues, developmental stages and physiological conditions, expression data of noncoding RNA genes is limited. Generation of large-scale datasets for both coding and noncoding RNAs allow for integrated analysis of mRNA-ncRNA interaction networks. In this project, we conducted companion sequencing of small RNAs in catfish after ESC and columnaris disease infection and hypoxia stress. Over 330 million 50-bp reads were generated, ~25 million reads for each sample. A total of 706 miRNA genes were

identified from the channel catfish genome using homology-based and sequencing-based approaches with the programs of MapMi and miRDeep2. Bioinformatic analysis identified 29, 17, and 15 differentially expressed miRNA genes after ESC infection, after columnaris infection and after hypoxia treatment, respectively, with distinct expression patterns. Correlated gene expression analysis of differentially expressed miRNAs revealed sets of negatively correlated expression of miRNAs and protein-coding genes, suggesting the potential target genes for the miRNAs. Potential mechanisms for the induced and correlated expression were investigated, providing insights into miRNA-mRNA interaction networks in response to bacterial infections and hypoxic stress. This work should be valuable to develop catfish genome expression Atlas for both coding and noncoding genes, and is important for the identification of biomarkers of bacterial diseases and hypoxia.

W048: Aquaculture

Genomics in Fish Breeding Programs for Developing Countries

John A.H. Benzie, WorldFish, Bayan Lepas, Malaysia

The majority of fish genetic improvement programs have focused to date on increased growth. This is particularly the case in developing countries where the penetration of genetically improved stocks is in any case limited. For example, worldwide the penetration of genetically improved stocks into production systems is still only 10-12% while in Africa it is a tenth of that. In circumstances where it is a challenge to provide basic infrastructure and inputs for aquaculture investment in more sophisticated approaches to genetic improvement programs are often questioned. At the same time however, there are demands for more resilient or robust fish that can perform in a variety of environments and for greater sustainability of approaches. The latter demands can only be met by application of genomic tools and the strategies for future research at WorldFish are anchored in the introduction of genomic tools not only in the development of breeding programs, but in the important task of identifying adoption and impact of improved strains. A case study in the history of genetic improvement in developing countries, and the changing priorities leading to more genomic applications, is Genetically Improved Farmed Tilapia (GIFT) originally developed by WorldFish and partner organizations. This example will be used to illustrate the challenges and opportunities for genomic research in international development.

W049: Aquaculture

Allelic-Imbalance Analysis in Pooled RNA-Seq Samples Identifies Muscle-Associated Genetic Markers in Rainbow Trout: Improved Bioinformatics Practices

Rafet Al-Tobasei, Middle Tennessee State University, Murfreesboro, TN

Coding SNPs (cSNPs) are likely to change the biological function of a protein. We have demonstrated that assessment of phenotype-associated allelic imbalances in pooled RNA-Seq samples is a fast means of identifying “large-effect” genetic markers. In this study, we compared two main variant calling bioinformatics pipelines, GATK and SAMtools, for discovery of cSNPs with allelic-imbalances that are associated with total-body weight, muscle yield, fat content, shear force and whiteness. Phenotypic data were collected from 500 fish representing 98 families (5 fish/family) of a growth-selected line. RNA-Seq was used to sequence the muscle transcriptome from 26 families showing divergent phenotypes (4 high- versus 4 low-ranked families/each trait). GATK detected 468,494 putative cSNPs: 6,228 showed allelic imbalances (>1.4 as an amplification and <0.7 as loss of heterozygosity). SAMtools detected 188,995 putative cSNPs; 943 had significant allelic imbalances between the low and high families. A total of 648 cSNPs with allelic imbalances were common between the two datasets, indicating significant differences in algorithms. A total of 6,966 non-redundant cSNPs were identified in the two datasets; 3,904 mapped to 1,606 protein-coding genes (with 13.5% non-synonymous cSNPs), and 1,279 mapped to 588 lncRNAs. Twenty cSNPs mapped to 6 genes showing differential phenotypic expression. Optimization of GATK and SAMtools to fit RNA-Seq pooled-sample analysis will be presented. These results will be used to build a cSNP-chip for genomic selection of muscle growth and quality traits in rainbow trout.

W050: Aquaculture

Role of Long Non-Coding RNAs in Bacterial Cold Water Disease Pathogenesis in Rainbow Trout

Bam D Paneru, Middle Tennessee State University, Murfreesboro, TN

Background: Bacterial cold water disease (BCWD) caused by *Flavobacterium psychrophilum* is one of the major causes of mortality in salmonids. Three genetic lines of rainbow trout designated as ARS-Fp-R (resistant), ARS-Fp-C (control) and ARS-Fp-S (susceptible) have significant differences in survival rate following *F. psychrophilum*infection. Previous studies have identified transcriptomic differences of immune-relevant protein-coding genes at basal and post-infection levels among these genetic lines.**Results:** Using an RNA-seq approach, we quantified differentially expressed long non-coding RNAs (lncRNAs) in response to *F. psychrophilum* challenge. Pairwise comparison between genetic lines and different stages of infection identified 705 differentially expressed lncRNAs. A positive correlation was observed between the number of the differentially regulated lncRNAs and that of the protein-coding genes following infection. Many of the differentially expressed lncRNAs were located near immune-related protein-coding loci. The expression patterns of several lncRNAs were either positively or negatively correlated with those of their neighboring and distant immune related protein-coding genes. The list includes complement components, suppressor of cytokine signaling, tumor necrosis factor receptor superfamily members, chemokines, and transcription factors involved in immunological responsiveness. Most of the protein-coding genes that correlated in expression with lncRNAs were transcription factors, activators, receptors, and molecules of signal transduction pathways related to the innate immune system. The correlated expression and genome-wide co-localization suggested that some of the lncRNAs may be involved in regulating immune-relevant protein-coding genes. **Conclusion:** This study provides the first evidence of correlated expression between lncRNAs and immune related protein-coding genes in an important aquaculture species.

W051: Aquaculture

Genotyping in Thousands By Sequencing (GT-seq): A Low Cost, High-Throughput, Targeted SNP Genotyping Method

Nathan Campbell, Columbia River Inter-Tribal Fish Commission, Hagerman, ID

GT-seq is a genotyping method which leverages large read numbers from Illumina sequencers to genotype hundreds of single nucleotide polymorphisms within pools of multiplex PCR amplicons generated from thousands of individual samples (Campbell *et al.* 2014). This method

produces genotypes that are 99.9% concordant to those produced using TaqMan™ assays at approximately 1/4th the cost. Since its development, GT-seq panels have been created for several species (Chinook salmon, coho salmon, sockeye salmon, rainbow trout, and pacific lamprey) and has become the preferred SNP genotyping method in our laboratory. New genotyping software allows genotypes and summary figures to be produced from a lane of raw sequencing data in under an hour using a desktop linux computer.

W052: Aquaculture

Development of the Catfish 690K SNP Arrays for Analysis of Quantitative Traits

Qifan Zeng, Qiang Fu, Shikai Liu and Yun Li, Auburn University; The Fish Molecular Genetics and Biotechnology Laboratory, Auburn, AL

Availability of genome-wide molecular markers is crucial for the analysis of quantitative traits. Single nucleotide polymorphisms (SNPs) are one of the most widely used molecular markers due to their ubiquitous distribution and analytical simplicity. In this work, we developed a high-density SNP genotyping array for channel catfish (*Ictalurus punctatus*) and blue catfish (*I. furcatus*). Whole genome sequencing, RNA-seq, and GBS data over 700 channel catfish and 200 blue catfish from previous experiments were used for SNP mining. Criteria-based quality filtration and Affymetrix *in-silico* probe converting test were used to select the final set of 690,662 high quality SNPs, comprising 581,002 channel catfish specific, 44,694 blue catfish specific, 19,124 inter-species, and 45,842 ubiquitous SNPs. A total of 48,434 strain-specific markers for 5 domestic and wild strains were also included on the array, which should be useful in study channel catfish domestication events. The selected SNPs were relatively evenly distributed on the genome reference, with 0.5-1.3 kb interquartile range of intervals. They cover 98.6% of the genome scaffolds and 93.9% of the BAC-based physical map contigs. The high coverage benefits the detection of linkage disequilibrium in further genome-wide association studies and fine-scale linkage mapping. To represent gene-associated regions well, 24,878 of 26,661 genes on the reference genome sequence were covered with at least one marker. 2,000 additional probes generated from non-polymorphic genomic regions were also included as negative controls. This is the highest density SNP array developed in aquaculture species, which should serve as valuable resources for further construction of genetic map, genome-wide association studies, and QTL analysis.

W053: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

Community Collaborations: Arabidopsis Informatics Consortium and Advancing Arabidopsis Research and Training

Joanna Friesner, UC Davis, Davis, CA

W054: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

A Tour of the Arabidopsis Information Portal

Agnes P Chan, J. Craig Venter Institute, Rockville, MD

Araport (https://www.araport.org) aims to to provide Arabidopsis and plant scientists direct access to a new generation web-based data platform. Users can browse and analyze a wide array of data through Araport. Users can also publish their own modules for sharing data with the community or building analysis workflows. In addition, Araport has assumed the responsibility for updating and revising genome annotation (Araport11) and the reference genome sequence.

In this workshop, we will demonstrate the powerful **ThaleMine** data warehouse for browsing Araport11 gene reports that are already integrated with genome annotation, coexpression, physical and genetic interactions, pathways, germplasm, etc., running gene list enrichment analysis, building data queries, exporting data tables, and saving/sharing work. We will also demonstrate the rich data content of the next generation genome browser **JBrowse** including the latest Araport11 gene structures, 1001 genomes variants, community data tracks, etc., and a simple way for users to upload their own data tracks for side-by-side viewing and sharing with collaborators. Finally, we will review a growing collection of data and analysis modules (**Science Apps**) that will serve as building blocks for creating discovery workflows.

Araport tutorials are available at, https://www.araport.org/tutorials.

Araport is supported by the NSF and the BBSRC.

W055: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

Module Development for Araport

Jason R. Miller, J. Craig Venter Institute, Rockville, MD

W056: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

User friendly tools for the Arabidopsis thaliana 1001 Genomes

Beth Rowan, Max Planck Institute for Developmental Biology, Tübingen, Germany

Arabidopsis thaliana is a not only an important model species for laboratory research, but also for the study of natural intraspecific genetic variation. Through the 1001 Genomes project, we have catalogued genetic variation in 1135 high-quality re-sequenced natural inbred lines representing the global population in its native Eurasian and North African range, and in recently colonized North America. Our data include over 10 million single nucleotide polymorphisms and 1.5 million small insertion-deletion polymorphisms (about one every 10 bp on average), which makes it the densest variant collection for any organism. To make this important resource of genomic variation more accessible and useful for the community, we have developed a suite of web-based interfaces for querying and analyzing the data. Users can subset the full genome variant call format or pseudogenome files by strain or locus, or find the most closely related strain in the 1001 genomes for a strain of interest. In this talk, I will demonstrate these and other tools in the development pipeline and discuss how they will be integrated into the Arabidopsis Information Portal.

W057: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

HRGRN: enabling graph search and integrative analysis of Arabidopsis signaling transduction, metabolism and gene regulation networks

Xinbin Dai[1], Jun Li[2], Tingsong Liu[1] and Patrick Xuechun Zhao[1], (1)The Samuel Roberts Noble Foundation, Ardmore, OK, (2)University of Texas MD Anderson Cancer Center, Houston, TX
The biological networks controlling plant signal transduction, metabolism and gene regulation are composed of not only genes, RNA, protein and compounds but also the complicated interactions among them. Yet, even in the most thoroughly studied model plant *Arabidopsis thaliana*, the knowledge regarding these interactions are scattered throughout literatures and various public databases. Thus, new scientific discovery by exploring these complex and heterogeneous data remains a challenge task for biologists.
We developed a graph-search empowered platform named HRGRN to search known and, more importantly, discover the novel relationships among genes in Arabidopsis biological networks. The HRGRN includes over 51,000 “nodes” that represent very large sets of genes, proteins, small RNAs, and compounds and approximately 150,000 “edges” that are classified into nine types of interactions (interactions between proteins, compounds and proteins, transcription factors (TFs) and their downstream target genes, small RNAs and their target genes, kinases and downstream target genes, transporters and substrates, substrate/product compounds and enzymes, as well as gene pairs with similar expression patterns to provide deep insight into gene-gene relationships) to comprehensively model and represent the complex interactions between nodes. . The HRGRN allows users to discover novel interactions between genes and/or pathways, and build sub-networks from user-specified seed nodes by searching the comprehensive collections of interactions stored in its back-end graph databases using graph traversal algorithms. The HRGRN database is freely available at http://plantgrn.noble.org/hrgrn/. Currently, we are collaborating the Araport team to develop REST-like web services and provide the HRGRN’s graph search functions to Araport system.

W058: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

PMR metabolomics and transcriptomics database and its RESTful web APIs: A data sharing resource

Manhoi Hur, Iowa State University, Ames, IA
PMR database is a community resource for deposition and analysis of metabolomics data and related transcriptomics data. PMR currently houses metabolomics data from over 25 species of eukaryotes. In this talk, we introduce PMRs RESTful web APIs for data sharing, and demonstrate its applications in research using Araport to provide Arabidopsis metabolomics data.

W059: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

Predicting differential intron retention with iDiffIR

Michael Hamilton, Colorado State University, Fort Collins, CO

W060: Arabidopsis Information Portal & Intl Arabidopsis Informatics Consortium

ProtAnnot: visualizing effects of alternative splicing and transcription on protein sequence and function in a genome browser

Ann Loraine[1], Tarun Mall[2], John Eckstein[2], David Norris[2] and Nowlan Freese[2], (1)UNC-Charlotte, Kannapolis, NC, (2)University of North Carolina Charlotte, Kannapolis, NC
One gene can produce multiple transcript variants encoding proteins with different functions. To facilitate visual analysis of transcript variants and elucidate gene function, we developed ProtAnnot, which shows protein annotations in the context of genomic sequence. ProtAnnot searches InterPro and displays profile matches (protein annotations) alongside gene models, exposing how alternative promoters, splicing, and 3' end processing add, remove, or remodel functional motifs. To draw attention to these effects, ProtAnnot color-codes exons by frame and displays a cityscape graphic summarizing exonic sequence at each position. These techniques make visual analysis of alternative transcripts faster and more convenient for biologists. ProtAnnot is a available as a plug-able App for Integrated Genome Browser, an open source desktop genome browser available from http://www.bioviz.org.

W061: Arthropod Genomics

All the Better to Eat You with: Identifying and Characterizing Digestive Cysteine Peptidases in *Tribolium castaneum*

Lindsey Perkin, USDA-ARS, Manhattan, KS
The red flour beetle, *Tribolium castaneum*, is a major agricultural pest responsible for considerable loss of stored grain and cereal products worldwide. *T. castaneum* larvae have a highly compartmentalized gut, with cysteine peptidases mostly in the acidic anterior part of the midgut which are critical to the early stages of food digestion. We associated specific cysteine peptidase genes with digestive functions for food processing based on a comparison of gene expression profiles in different developmental stages (egg, larvae, pupae, and adult). From this data we created a new model of food digestion in *T. castaneum*. Based on this model, we selected a major digestive cysteine peptidase to target via RNAi. By monitoring gene expression with transcriptome analysis, we determined knockdown specificity, off-target effects, and digestive compensation by other digestive peptidases. This data contributes new information on digestion in an agricultural pest and will be helpful in developing targeted pest control strategies for tenebrionid insects.

W062: Arthropod Genomics

Genome-Wide SNPs Decipher Global Incursion Pathways in the *Bemisia tabaci* Species Complex

Samia Elfekih[1], Paul Etter[2], Eric Johnson[2], Karl Gordon[3], Paul De Barro[4] and Wee Tek Tay[5], (1)CSIRO, Canberra, Australia, (2)University of Oregon, Eugene, OR, (3)CSIRO Biosecurity flagship, Australia, canberrr, Australia, (4)CSIRO biosecurity flagship, Australia, Brisbane, Australia, (5)CSIRO biosecurity flagship, Australia, Canberra, Australia
The sweetpotato whitefly *Bemisia tabaci* is a complex of morphologically indistinguishable species and is one of the most destructive insect pests worldwide. This pest is multivoltine, highly polyphagous and a vector of several plant viruses. Previous studies on *B. tabaci* relied on single partial mitochondrial DNA gene and limited nuclear DNA markers to infer species status and populations interconnectedness. However, the limited information available on their genetic make-up affected our ability to understand and trace their global routes of invasion. Here, we conducted a genome-wide SNPs analysis to pinpoint the global pathways for the two most invasive species (MED and MEAM1) in the *B.*

tabaci complex. Using the Nextera-tagmented reductively amplified DNA (nextRAD) approach, which is a PCR-based protocol that mimics RAD-sequencing without restriction enzymes, we successfully obtained a total number of 3,331 SNPs present in all individuals and distributed throughout the genome from single whitefly individual samples.

The analysis was conducted on 7 and 8 populations respectively in MED and MEAM1 species. These populations provide a good representation of the geographical range where the species occur. The robust nextRAD dataset (no missing data), allowed us to infer the possible migration routes of each species starting from sub-Saharan Africa to the East Mediterranean region and the rest of the globe. Here, we demonstrate that genome-wide SNPs can be used to document the history of admixture and trace the movements of invasive species in order to enhance biosecurity threat identification.

W063: Arthropod Genomics

Draft Genome of *Candidatus* Liberibacter solanacearum Haplotype C (LsoC) from Carrot Psyllids Removed from Symptomatic Carrots in Norway

Kent F. McCue, USDA Agricultural Research Service, Albany, CA, Gerard R. Lazo, USDA Agricultural Research Service, WRRC, Albany, CA and Joseph E. Munyaneza, USDA Agricultural Research Service, Wapato, WA

The fastidious prokaryote *Candidatus* Liberibacter solanacearum (Lso), transmitted by the tomato potato psyllid (*Bactericera cockerelli*), is associated with the Zebra Chip disease of potato. Plants infected with Liberibacter may experience significant yield losses and these plants also serve as potential reservoirs for Liberibacter to spread to other psyllid hosts with alternate crop feeding preferences, potentially introducing new epidemiological focal points among crops. New associations between Liberibacter species and crop plants have been detected in different parts of the world, presenting concerns about the potential roles of these strains in causing disease. Carrots (*Daucus carota*) showing damage from the feeding of carrot psyllids (*Trioza apicalis*) were reported to be infected with a new haplotype of Lso designated LsoC. We sequenced and analyzed genomic contigs of DNA isolated from psyllids collected in Norway that tested positive for the presence of Lso. Using Illumina high-throughput short-sequence methods, we generated contigs and assembled a draft genome for LsoC. Differences between the LsoC were identified and contrasted with the LsoB haplotype from the United States. These differences can be used to understand the evolutionary divergence of haplotypes and factors involved in psyllid-vector/plant-host interactions.

W064: Arthropod Genomics

Sialotranscriptomics of *Rhipicephalus appendiculatus* Male and Female Ticks

Minique H de Castro[1], Daniel de Klerk[1], Ronel Pienaar[1], Abdalla Latif[1], Jasper Rees[2] and Ben J Mans[1], (1)Agricultural Research Council, Onderstepoort, South Africa, (2)Agricultural Research Council, Pretoria, South Africa

Ticks are blood-feeding ectoparasites that vector a variety of human and veterinary diseases worldwide. By injecting secretory proteins into the host they create a stable feeding site to feed unnoticed for extended periods of time. Tick secretory proteins modulate the host's haemostasis, inflammation and immune responses and are attractive vaccine candidates. Many of these proteins are uncharacterised due to limited sequence availability for ticks and other arthropods, but recent advances in next generation sequencing technologies have made sequencing of non-model organisms such as arthropods more affordable. To this end, we sequenced adult male and female *Rhipicephalus appendiculatus* salivary glands during feeding. Illumina MiSeq and HiSeq 2000 sequencing instruments were used to generate nearly 430 million paired end reads. A sialotranscriptome of 21 410 transcripts was *de novo* assembled from the trimmed reads and translated into 12 761 non-redundant proteins. Transcripts annotated as secretory proteins represented the smallest class of proteins in the transcriptome (17%) but accounted for most of the expression (63%). Of the secretory protein class expression, transcripts of the Glycine rich superfamily contributed 66% and transcripts of the Lipocalin family, 12%. Gender-skewed transcript expression was observed for a number of secretory protein families. Differential expression analysis identified 2346 male and 1758 female up regulated transcripts, suggestive of varying blood-feeding mechanisms employed between male and female ticks. This sialotranscriptome greatly improves on the sequence information available for *R. appendiculatus*, will improve our understanding of tick feeding biology and will be a valuable resource for future vaccine candidate selection.

W065: Arthropod Genomics

Tomato Infected with the Semipersistent Tomato Chlorosis Virus Results in Temporally Altered Gene Expression in *Bemisia tabaci*

NK Kaur, USDA-ARS, Salinas, CA

Whitefly transmission of semi-persistent viruses requires specific interactions between the virus and its vector, but little information exists on how host plant infection by semipersistent viruses influences the whitefly or its ability to acquire and retain a plant virus. In order to determine changes in global transcriptional gene regulation in whiteflies resulting from feeding on plants infected with a semipersistent virus, paired-end 151-bp RNA-Seq was performed on whitefly (*Bemisia tabaci* MEAM1) after 24, 48, and 72 hours of feeding on *Tomato chlorosis virus* (ToCV)-infected and healthy tomato in three biologically replicated tests. Among a total of 15,668 genes predicted in the whitefly genome, we identified significantly differentially expressed genes in whiteflies fed on ToCV-infected tomato: 447↑543↓ (at 24 h), 4↑ 7↓ (at 48 h), and 50↑160↓ (at 72h), by comparing differential expression based on fold changes no less than 1.5. This indicates temporally influenced gene expression in whiteflies due to the presence of ToCV in host plant phloem, and that gene expression is influenced within the first 24 h of feeding, with fewer and different genes activated after 72 h feeding, corresponding with the time at which ToCV retention in the vector ceases. Analysis of global differential expression revealed up-regulation of genes related to *cathepsin, glucose transporter* family members, uric acid pathway and up and down regulation of *alpha-glucosidases* in whiteflies fed on ToCV infected tomatoes. Results may reveal pathways associated with whitefly-species specific virus transmission, and will serve as a valuable resource toward effective whitefly and virus control.

W066: Arthropod Genomics

Using NextRAD Sequencing to Infer Fine-Scale Movement of Insects

Daisy (Zhen) Fu, Washington State University, Pullman, WA

Winged herbivorous insects can readily move between habitats and/or host plants. Analysis of genetic variation has been one tool to understand insect dispersal patterns. Existing approaches have generally been most effective when looking at broad-scale movement patterns, where genetic variation at a relatively coarse scale is informative. Here, we explore the use of Nextera-tagmented reductively-amplified DNA ("NextRAD") to explore relatively fine-scale movement patterns of the potato psyllid, *Bactericera cockerelli*. The potato psyllid is of fundamental interest as a polyphagous herbivore that disperses broadly to reach different host-plant species, and is of applied interest as the vector of zebra chip disease of cultivated potato (*Solanum tuberosum*). We identified 8,443 single nucleotide polymorphism (SNPs) of potato psyllids separated spatiotemporally in potato fields or on a key non-crop overwintering host, bittersweet nightshade (*S. dulcumara*), across an agricultural landscape. A subset of the potato psyllids on potato exhibited close genetic similarity to insects in nearby nightshade patches, inferring movement of the insects between these two host plants. However, a second subset of potato-collected psyllids was genetically distinct from those collected on nightshade; this suggests that another, currently un-recognized host plant could be contributing to psyllid populations in potato. Altogether, our results provide evidence that NextRAD sequencing is an effective means to delineate complex patterns of insect movement across landscapes. This approach may have broad utility for the many systems where plant or animal pathogens are vectored by small, highly-mobile insects.

W067: Arthropod Genomics

Investigating the Role of *Rickettsia* in *Tomato Yellow Leaf Curl Virus* Interactions with its Whitefly Vector *Bemisia tabaci*

Murad Ghanim[1], Michelle Cilia[2] and Adi Kliot[1], (1)Volcani Center, Bet Dagan, Israel, (2)Cornell University, Ithaca, NY

Rickettsia, a secondary bacterial symbiont that resides within the whitefly *Bemisia tabaci* has been recently implicated in *Tomato yellow leaf curl virus* (TYLCV)-*B. tabaci* interactions. Unlike the rest of the secondary symbionts that infect *B. tabaci* and reside in a specialized organ, *Rickettsia* occupies most of the body cavity of the insect, including organs implicated in virus transmission. Using Fluorescence in-situ hybridization (FISH) and qPCR we have shown that whiteflies infected with *Rickettsia* acquire and transmit more virus particles than those uninfected with *Rickettsia*; both of the same genetic background. New experiments suggest that *B. tabaci* adults infected with *Rickettsia* are more attracted to TYLCV-infected tomato plants. In addition, adults infected with both TYLCV and *Rickettsia* has improved fecundity but show a negative effect on the fitness of their offspring. A transcriptomic analysis performed on *Rickettsia*-infected and uninfected *B. tabaci* adults before and after 48 h TYLCV acquisition showed that TYLCV induces dramatic gene expression changes in both *Rickettsia*-infected and uninfected whiteflies, however, while TYLCV acquisition causes up-regulation of numerous genes in the *Rickettsia* uninfected population, it causes massive down-regulation in the *Rickettsia* infected ones. The results showed induction of several immunity related genes even prior to TYLCV acquisition, possibly as a response to infection with *Rickettsia*. A proteomic analysis including 9 populations with different TYLCV transmission abilities indicated the involvement of *Rickettsia* in improved TYLCV transmission ability. These studies set the foundation for understanding the role of *Rickettsia* in many aspects of *B. tabaci* interactions with its environment.

W068: Avian Genomics - Going Wild!

Generating a Bird Genome Resource: Insights into the Avian Tree of Life, Complex Traits, and Genome Evolution

Erich Jarvis, Duke University Medical Center, Durham, NC

W069: Avian Genomics - Going Wild!

Non-Model Avian Genomics Fit for Purpose - the End of the Model Organism?

David W. Burt, Roslin Institute Univ of Edinburgh, Edinburgh, United Kingdom

W070: Avian Genomics - Going Wild!

Comparative Population Genomics of Birds

Robb T. Brumfield, Louisiana State University, Baton Rouge, LA

W071: Avian Genomics - Going Wild!

The Challenges of Repetitive DNA in Avian Comparative Genomics

Alexander Suh, Dept. of Evolutionary Biology (EBC), Uppsala University, Uppsala, Sweden

Repetitive DNA sequences are ubiquitous components of genomes and distributed either as tandem repeats (e.g., centromeres, telomeres, microsatellites) or as interspersed repeats (e.g., transposons, endogenous viruses). The recent "big bang" in avian genomics has made a plethora of genome-scale information readily available, however, the study of the impact of repetitive elements on the evolution of birds remains challenging using current techniques for sequencing and genome assembly. Here, I will discuss recent advances in the study of repetitive DNA across the evolution of birds and their closest relatives. Although much of avian repetitive DNA still constitutes genomic "dark matter", there is a growing body of evidence showing long-lasting associations of birds with specific groups of transposons and viruses. Finally, I will demonstrate that state-of-the art approaches are capable of successfully sequencing, assembling, and characterizing "dark matter" in avian genomes.

W072: Avian Genomics - Going Wild!

The Role of Hybridization in the Evolutionary History of Geese

Jente Ottenburghs, Wageningen University, Wageningen, Netherlands

W073: Avian Genomics - Going Wild!

Genomics and Speciation: The New World Mallard Complex

Philip Lavretsky, University of Miami-Florida, Coral Gables, FL

The New World (NW) mallard group includes the dichromatic mallard (*Anas platyrhynchos*), and four monochromatic taxa; American black duck, (*A. rubripes*), Mexican duck (*A [p.] diazi*), and two subspecies of mottled ducks (Florida, *A. fulvigula fulvigula*; and West Gulf Coast. *A. f.*

maculata). Although all NW taxa are phenotypically diagnosable, resolving their taxonomic relationships has been challenging due to genomic similarities attributable to their recent ancestry and/or hybridization. Using ddRAD-seq methods, we sequenced 3,029 autosomal and 198 Z-chromosome markers from a total of 166 samples (24–43 per taxon). For both marker-types, the monochromatic taxa each clustered into independent groups; however, American black ducks were largely indistinguishable from mallards. Under a neutral scenario in which genetic divergence is driven by genetic drift, the expected ratio for Z:autosomal differentiation is < 1.33, because Z loci have 0.75 the effective population size of autosomal loci. Comparing mallards to each monochromatic taxa recovered elevated Z divergence, with Z:autosomal ratios ranging between 4 and 6.5, and evidence of divergent selection acting on 2-4% of Z-linked markers, but < 0.5% of autosomal markers. Furthermore, aligning markers along the Z chromosome revealed a region of elevated differentiation that was shared between all mallard-monochromatic comparisons; this region was less-differentiated or absent for pair-wise comparisons of monochromatic taxa. These results suggest that the Z-chromosome is at a later stage of divergence between mallards and monochromatic taxa and that this differentiation is likely driven by selection acting on traits that are Z-linked and derived in mallards. Overall, ddRAD-seq markers provided high resolution regarding the evolutionary history of the NW mallard clade. Furthermore, they revealed strong support for phylogenetic relationships within this group, suggesting that black ducks and mallards are sister species, this pair is sister to Mexican ducks, and the two mottled duck populations are a separate lineage.

W074: Banana Genomics

A Metabolomics Approach to the Assessment of Banana Diversity and Traits

Margit Drapal[1], Elisabete Carvalho[2], Ines Van den houwe[3], Mathieu Rouard[4], Julie Sardos[4], Delphine Amah[5], Rony Swennen[6], Nicolas Roux[4] and Paul D. Fraser[7], (1)Royal Holloway University of London, London, United Kingdom, (2)Royal Holloway University of London, Surrey, United Kingdom, (3)Bioversity International Transit Center, Heverlee, Belgium, (4)Bioversity International, Montpellier, France, (5)IITA, Ibadan, Nigeria, (6)KU Leuven, Leuven, Belgium, (7)Royal Holloway University of London, Egham, United Kingdom

Banana (*Musa*) is one of the most important economic and staple crops in the world. The majority of edible cultivated banana species arises from two species of the Eumusa group, *Musa acuminata* (A genome) and *Musa balbisiana*(B genome). In order to assess the biochemical diversity that exists in our banana germplasm collections multi-platform metabolomics platforms has been established for banana. These include LC-MS in untargeted and targeted mode, GC-MS based metabolite profiling and targeted UPLC-PDA for compounds such as carotenoids where MS ionisation is poor.

Metabolomic finger printing and complementary targeted analysis has been performed on *in vitro* vegetative material for 20 diverse Musa accessions, including diploid varieties, wild *Musa acuminata* and *Musa balbisiana* as well as different triploids and distant wild species (*Musa ornata*). The data allowed the separation of the genotypes on the basis of genotypes and differentiating metabolites identified between accessions. Comparisons with field grown material was carried out in selected cases and clear correlation was observed including the potential to predict fruit phenotypes on vegetative profiles. These robust techniques can now be utilised in combination with of omic approaches to characterise consumer and agronomic traits within breeding populations.

W075: Banana Genomics

One Hundred and Six Diploids to Unravel the Genetics of Traits in Banana: A Panel for Genome-Wide Association Study and its Application to the Seedless Phenotype

Julie Sardos[1], Mathieu Rouard[1], Yann Hueber[1], Alberto Cenci[1], Katie Hyma[2], Ines Van den houwe[3], Eva Hřibová[4], Brigitte Courtois[5] and Nicolas Roux[1], (1)Bioversity International, Montpellier, France, (2)Genomic Diversity Facility, Cornell University, Ithaca, NY, (3)Bioversity International Transit Center, Heverlee, Belgium, (4)Institute of Experimental Botany, Prague, Czech Republic, (5)CIRAD, Montpellier Cedex 5, France

Banana is a fruit crop with a complex diversity pattern resulting from a complex domestication scheme. Due to the availability of plant genome sequences and to the accessibility of next-generation genotyping technologies, Genome-Wide Association Studies (GWAS) have been increasingly performed in crop plants as a start to resolve genetic architecture of traits. The GWAS method was developed to perform association studies between phenotypes such as diseases and genotypes in Humans and is now successfully applied to many plants and animals to support the breeding process. However, there are prerequisites for such methods as it has been designed for populations of diploids organisms which follow the mendelian genetics model (i.e. “infinite” population and panmixia). In banana, the most popular cultivars are triploids, often hybrids between different species, and due to the absence of seeds in the fruit, a wide amount of the diversity observed ensues from the clonal diversification of a few initial genotypes. Therefore, the application of such approach to the crop is challenging and innovative. We selected a set of 106 diploid accessions with pure *M. acuminata* background avoiding clone duplicates and generated the appropriate molecular markers to support GWAS for any given trait. Finally, we validated the approach on the major domestication syndrome in banana, the seedless phenotype, and identified six candidate regions in which are located two strong candidate genes for female sterility.

W076: Banana Genomics

A Fungal Genomics Perspective on the Panama Disease Threat of Banana

Gert HJ Kema, Wageningen University and Research Center, Wageningen, Netherlands

Knowledge on population diversity is key to any potential disease management strategy. We have used overall Foc genotyping to investigate the current dissemination of the causal agent of Panama disease, Fusarium oxysporum f.sp. cubense, and determined that - in spite of massive global diversity - just one clone threatens Cavendish bananas. This clone is also pathogenic on a wide variety of banana germplasm and hence discovering and deploying resistance is urgent. The latest data will be presented.

W077: Banana Genomics

Transcriptomic Profiling in *Musa*: A Look into Processes Affected by Mild Osmotic Stress in the Root Tip

Jassmine Zorrilla[1], Mathieu Rouard[2], Alberto Cenci[2], Ewaut Kissel[1], Hien Do[1], Emeric Dubois[3], Sabine Nidelet[3], Nicolas Roux[2], Rony Swennen[1] and Sebastien C Carpentier[4], (1)KU Leuven, Leuven, Belgium, (2)Bioversity International, Montpellier, France, (3)MGX-Montpellier GenomiX, Montpellier, France, (4)SYBIOMA, Leuven, Belgium
Drought stress is one of the major abiotic factors limiting banana (*Musa*) production. Even mild-drought conditions are responsible for considerable yield losses. We performed large-scale transcriptome sequencing using Illumina technology on root tissue of three triploid genotypes representing well known cultivars and focused on the identification of genes with an altered expression pattern under mild osmotic stress (3 days after 5% PEG treatment). In total, 18 cDNA libraries were sequenced producing around 568 million high quality reads, of which 70-84% were mapped to the diploid reference genome (D'Hont et al., 2012). Through uni-/multivariate statistics, 92 genes were commonly identified as differentially expressed in the three genotypes. Using our in house workflow to analyze GO enriched and underlying biochemical pathways, we present a panorama of the general processes affected by mild osmotic stress in the root tip, although we observe a bias towards glycolysis and fermentation. We hypothesize that in this fast growing and oxygen demanding tissue, mild osmotic stress leads to a lower energy level, which induces a metabolic shift towards (i) a higher oxidative respiration, (ii) alternative respiration and (iii) fermentation. To validate the mRNA-seq results, a subset of twenty up-regulated genes were further analyzed at three different time points (6 hours, 3 days and 7 days) in an independent PEG experiment. Overall, the identification and annotation of this set of genes constitutes a step ahead to understand the complex network of root responses to osmotic (drought) stress.
References:
D'Hont A., et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature, 2012. 488: p. 213-217.

W078: Banana Genomics

Genomic Breeding Approaches for East African Bananas

Moses Nyine[1], Brigitte Uwimana[1], Rony Swennen[2], Michael Batte[1], **Allan Brown**[3], Eva Hřibová[4] and Jaroslav Dolezel[5], (1)International Institute of Tropical Agriculture, Kampala, Uganda, (2)KU Leuven, Leuven, Belgium, (3)International Institute of Tropical Agriculture, Arusha, Tanzania, (4)Institute of Experimental Botany, Prague, Czech Republic, (5)Institute of Experimental Botany, Olomouc, Czech Republic
The polyploidy nature of banana is a limiting factor in the implementation of strategies such as marker assisted selection (MAS) or genome wide association mapping (GWAS). The triploid nature of cultivated varieties complicates conventional breeding strategies and improved varieties can take up to 20 years before they can be released to the public, which necessitates the use of efficient molecular tools to more rapidly respond to abiotic and biotic stresses and to address the needs of growers and consumers. In addition, the high cost of phenotyping perennial large-stature plants such as banana, and the rapidly decreasing cost of genotyping, makes the use of predictive genomic selection models using single nucleotide polymorphism (SNP) markers extremely attractive to banana breeders. A Genomic Selection (GS) training population consisting of 307 banana genotypes was developed for initial analysis with ploidy levels of the plant material ranging from diploids to tetraploids. Plants were genotyped using the genotyping by sequencing (GBS) approach (Elshire et al., 2011) with PstI as the sole restriction enzyme. Sequence data was processed through a bioinformatics workflow and single nucleotide polymorphisms (SNPs) were called using the genomic analysis tool kit (GATK). Data was filtered for quality and for loci with >50% missing data. Phenotypic data for 25 traits are being collected from two locations since 2012. Yield-related traits (fruit pulp diameter, bunch weight, number of suckers, etc.) are collected at flowering and harvest Analysis of GBS data resulted in 11201 SNP loci. The results of multiple prediction models are discussed and compared.

W079: Banana Genomics

A Pipeline for Generating Mutant Bananas Resistant to Fusarium Wilt TR4

Bradley J. Till[1], Joanna Jankowicz-Cieslak[1], Souleymane Bado[1], Anna Sochacka[1], Sneha Datta[1], Anne Davson[2], Chih-Ping Chao[3], Shih-Hung Huang[3] and Altus Viljoen[4], (1)International Atomic Energy Agency, Vienna, Austria, (2)Du Roi Laboratory, Letsitele, South Africa, (3)Taiwan Banana Research Institute, Pingtung, Taiwan, (4)Stellenbosch University, Stellenbosch, South Africa
Triploid bananas are sterile, parthenocarpic, and an obligate vegetatively propagated crop. Lack of meiosis hampers traditional breeding. Exported Cavendish bananas are clones and susceptible to diseases including Fusarium wilt caused by tropical race four (TR4). In recent years TR4 has been identified in nine countries suggesting that it is spreading geographically and threatening global banana production. Work with banana TILLING showed that mutations can be induced in the genome at a high density and chimeric sectors rapidly dissolved. This suggests that a mutation breeding approach is suitable to develop disease resistant bananas. For success, a large mutant population must be generated and efficiently screened to recover rare improved plants. Each step of the breeding pipeline (clonal production, mutagenesis and disease screening) represents a bottleneck that is difficult to overcome in a single laboratory. We are developing a large mutant population of approximately 10,000 lines by treatment of shoot tips with gamma irradiation, and modifying sequencing techniques to recover large genomic indels caused by the treatment. In parallel to this, the IAEA is developing a multi-laboratory Coordinated Research Project with a focus on banana disease resistance.

W080: BBSRC/NSF/ERA-CAPS - Challenges and Opportunities in Plant Science Data Management

Standards and Infrastructure for Diverse, Dispersed Data

Paul J. Kersey, EMBL - The European Bioinformatics Institute, Cambridge, United Kingdom
With next-generation sequencing, proteomic and metabolomic technologies, and greenhouse and field-based high-throughput phenotyping all increasingly common, a growing diversity of biological data is becoming available for plant species. While some of these data (e.g. nucleotide sequence data) are well-archived in international public repositories, other data types suffer from limited public availability, irregular data formatting and description, and dispersal over many resources. In this talk I will describe a number of efforts for addressing these problems, developed in the context of the transPLANT and EXCELERATE projects. These include the development of MIAPPE, a set of minimal information standards for plant phenotyping experiments; the development of a distributed approach for the integrated querying of dispersed, diverse data; a new framework for discovery and retrieval of plant phenotypic data; and a wider vision for how different resources (with particular specialities) might combine to present users with a single virtual resource. Although some of the challenges faced by the plant science

community are specific, others are shared with other biological researchers, and I will conclude by showing how these initiatives fit within the broader context of the ELIXIR life science data infrastructure.

W081: BBSRC/NSF/ERA-CAPS - Challenges and Opportunities in Plant Science Data Management
A pragmatic path forward for integrating phenotype and trait data using ontologies
Ramona Walls, The iPlant Collaborative, Tucson, AZ
Phenotypes are hard to deal with. Unlike genotypes, which can be described using sequences of characters, phenotypic data are conditional and non-standardized and therefore normally described using text. Even quantitative phenotype data (e.g., plant height in cm) require additional explanation such as how and when they measured to be understood fully. With the growth of new automated phenotyping and trait measurement methods, we are in desperate need of scalable, standardizable methods for annotating and storing phenotype and trait data, if we ever want that data to be discoverable, computable, and re-usable. Phenotype and trait ontologies (well-defined sets of terms with logical relationships linking them) hold the promise of standardizing phenotype data, but only a very small percentage of researchers use ontology terms to describe their data. In this talk, I will describe some recent efforts that make use of ontologies for describing phenotypes and lay out a roadmap for how to build on this work for more scalable data systems. Topics will include cross-species comparisons of mutant phenotypes, managing biodiversity trait data, and linking ontologies to metadata.

W082: BBSRC/NSF/ERA-CAPS - Challenges and Opportunities in Plant Science Data Management
Ontologies a language for data integration and reuse
Georgios Gkoutos, Computer Science Department, Aberystwyth University, Aberystwyth, United Kingdom

W083: BBSRC/NSF/ERA-CAPS - Challenges and Opportunities in Plant Science Data Management
Infrastructure and standards for publishing research data in the plant sciences
Robert Davey, The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland
The aim of open science is to make scientific research accessible, facilitating experimental reproducibility and transparency. Mechanisms exist for preserving and publishing research objects in plant science within the “omics” fields. However, researchers are often hindered by:
(i) complicated and time-consuming procedures for repository deposition;
(ii) a lack of interoperability between disparate information sources and mechanisms;
(iii) sub-optimal search and retrieval facilities across data repositories; (iv) a lack of public awareness of existing services.
To address these issues, we are developing COPO (Collaborative Open Plant Omics), a brokering service which enables aggregation and publishing of research outputs by plant scientists, and provides access to data repository services across disparate sources of information via web interfaces. COPO comprises a novel intuitive wizard-enabled web front-end, data grid infrastructure, and a number of Application Programming Interfaces (APIs), which facilitate the creation and management of logical profiles containing heterogeneous but related research objects representing a body of work. COPO leverages the Investigation Study/Assay (ISA) formats and ISA software suite (http://www.isa-tools.org) to enable experimental metadata attribution and conversion between metadata formats.
As research objects annotated and deposited through COPO will be accompanied by rich metadata implemented in JSON-LD (http://www.json-ld.org), cutting edge technologies such as MongoDB (http://www.mongodb.org) and neo4J (http://www.neo4j.org) are being used to create a web of linked semantic knowledge, allowing for user customised suggestions for future avenues of investigation.

W084: BBSRC/NSF/ERA-CAPS - Challenges and Opportunities in Plant Science Data Management
Data Processing and Analysis - Challenges for HT Phenotyping Data
Jesse Poland, Kansas State University, Manhattan, KS

W085: Bioenergy Grass Genomics
Exploiting Sorghum Diversity for Bioenergy Traits
Patrick J. Brown, University of Illinois, Urbana-Champaign, Urbana, IL
Sorghum is a resilient cereal that can be used to produce grain, sugar, and biomass. To fully exploit sorghum diversity and accelerate genetic improvement for different end uses, we need to understand which alleles are critical for distinguishing grain, forage, sweet, and biomass sorghum ideotypes. To address this need, we performed genome-wide association studies separately in panels of sweet sorghum and biomass sorghum. We find that different alleles at the sorghum *Dry Stalk (D)* locus increase yield in sweet sorghum and biomass sorghum. In sweet sorghum, green juicy midribs *(dd)* are associated with higher juice yield and higher sugar yield. In biomass sorghum, white dry midribs *(D-)* are associated with lower moisture, lower stalk lodging, and higher biomass yield. Analysis of *D* locus NILs confirms these phenotypic effects and suggests that the *D* locus affects vascular development. Certain crosses between green midrib inbreds give rise to white midrib segregants, indicating that multiple loci condition the green midrib phenotype in sorghum.

W086: Bioenergy Grass Genomics
Physiological and Molecular Analysis of Drought Stress Responses in Miscanthus: Identifying Important Traits for Biofuel Production from Analysis of Networks
Marta Malinowska[1], Vera Vendramin[2], Simone Scalabrin[2], Sabine Schnabel[3], Iain Donnison[1] and Paul R. H. Robson[1,4], (1)Aberystwyth University, Aberystwyth, United Kingdom, (2)IGA Technology Services, Udine, Italy, (3)Biometris, Wageningen UR Plant Science Group, Wageningen, Netherlands, (4)IBERS - Aberystwyth University, Aberystwyth, United Kingdom
Miscanthus is a genus of C4 perennial grasses of great interest for the production of biorenewable energy and chemicals because of high biomass potential even in temperate regions. Yield is strongly linked to water availability and many sites across Europe where irradiation and temperature are favourable for *Miscanthus* cultivation have limited water supply. The aim of the research is to investigate in detail the response of diverse

genotypes to reduced water availability, including the hybrid *M. x giganteus* that is grown for the majority of the crop and two exemplar genotypes from each of the species groups, *M. sacchariflorus* and *M. sinensis*, that hybridized to form *M. x giganteus.*
The five *Miscanthus* genotypes were subjected to two treatments, well-watered and mild drought, for 32 days under glasshouse conditions. Morphological and physiological measurements were made including growth rate, photosynthetic performance and water use efficiency. Plants were sampled at two time points to identify early and late responses to drought. Analyses included RNA sequencing for gene expression profiling and for compositional analysis of biomass (lignin, cellulose). The glasshouse measurements showed treatment affects differed between genotypes. From 116343 identified genes, 1219 were differentially expressed between control and drought treatments across all genotypes. Additionally 61 genes were found to be differentially expressed between the genotypes under drought compared to controlled conditions.
By comparing molecular, physiological and compositional responses to mild drought, we hope to identify pathways responsible for drought resistance and sensitivity and improve selections for plants with good biomass production under water stress.

W087: Bioenergy Grass Genomics

Genetic Resources from *Dactylis glomerata* and *Phalaris arundinacea* Collected on Marginal Land Sites in Europe

Susanne Barth[1], Trevor Roland Hodkinson[2], Paul Cormican[3], Ruby Prickett[2] and Manfred Klaas[4], (1)Teagasc, Carlow, Ireland, (2)Trinity College Dublin, Dublin 2, Ireland, (3)Teagasc Animal and Bioscience Research Department, Dunsany, Ireland, (4)Teagasc Crops Environment and Land Use Programme, Carlow, Ireland

Dactylis glomerata and Phalaris arundinacea are two C3 forage grasses with a circumboreal distribution with a potential for biomass applications. We assessed the genetic diversity, and thus the potential for breeding improvements and adaptation to different climatic and soil conditions in the Northern and central European context, and undertook collections for both species in five northern European countries along West/East and North/ South clines. The accessions were genotyped by next generation sequencing to obtain a large number of markers without having to rely on existing genomic resources.

W088: Bioenergy Grass Genomics

Accuracy of Genomic Prediction in Switchgrass Improved by Accounting for Linkage Disequilibrium

Guillaume P. Ramstein, Department of Agronomy, University of Wisconsin - Madison, Madison, WI
Authors: Guillaume P. Ramstein, Joseph Evans, Shawn M. Kaeppler, Robert B. Mitchell, Kenneth P. Vogel, C. Robin Buell, and Michael D. Casler

Switchgrass is a relatively high-yielding and environmentally sustainable biomass crop, but further genetic gains in biomass yield must be achieved to make it an economically viable bioenergy feedstock. Genomic selection is an attractive technology to generate rapid genetic gains in switchgrass and meet the goals of a substantial displacement of petroleum use with biofuels in the near future. In this study, we empirically assessed prediction procedures for genomic selection in two different populations consisting of 137 and 110 half-sib families of switchgrass, tested in two locations in the United States for three agronomic traits: dry matter yield, plant height and heading date. Marker data was produced for the families' parents by exome capture sequencing, generating up to 108,077 polymorphic markers with available genomic location and annotation information. We evaluated prediction procedures that varied not only by learning schemes and prediction models, but also by the way the data was preprocessed to account for redundancy in marker information. More complex genomic prediction procedures were generally not significantly more accurate than the simplest procedure, likely due to limited population sizes. Nevertheless, a highly significant gain in prediction accuracy was achieved by transforming the marker data through a marker correlation matrix. Our results suggest that marker-data transformations and, more generally, the accounting of linkage disequilibrium among markers, offer valuable opportunities for improving prediction procedures in genomic selection. Some of the achieved prediction accuracies should motivate implementation of genomic selection in switchgrass breeding programs.

W089: Bioenergy Grass Genomics

Genetic Controls of Biomass Increase in Sugarcane by Association with Beneficial Nitrogen-Fixing Bacteria

Adriana Hemerly, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
Carvalho, TLG[1]; Ballesteros, HGF[1]; Mota Filho, JP[1]; Hardoim, P[1].; Fernandes T.M.D[1]; Mendes, J.L.M. [1]; Atella, M. [1]; Baldani, JI[2]; Ferreira PCG[1], Hemerly, AS[1]. hemerly@bioqmed.ufrj.br
[1] Laboratório de Biologia Molecular de Plantas, Instituto de Bioquímica Médica Leopoldo de Meis, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, 21.941-590, Rio de Janeiro, RJ, Brazil; [2]Embrapa Agrobiologia, BR465, Km47, 23851-970, Seropédica, RJ, Brazil.

Sugarcane is the most important bioenergy grass in Brazil, being successfully used for first generation ethanol production from sugar. More recently, energy cane genotypes are being created in Brazil as an alternative for second generation ethanol production from lignocellulosic biomass. The success of the biofuel programs led to an increase in the demand of this crop productivity in the country for the next years. An important biotechnological challenge of this century is to develop tools to apply for a sustainable agriculture that would increase productivity using less fertilizers, pesticides, water and cultivated area. The associations that occur between sugarcane and other grasses with nitrogen-fixing associative and endophytic bacteria have raised a large interest in their use in agriculture, in view of the positive effects on biomass, productivity and tolerance against stresses. The promotion of plant growth by the endophytic nitrogen-fixing bacteria might be mediated by providing nitrogen trough Biological Nitrogen Fixation (BNF) and hormones. Inoculants of associative and endohytic diazotrophic bacteria had been shown to lead to positive results on sugarcane yields, which are dependent on the plant genotype and soil conditions. Our group has been studying plant genes involved in the establishment of a beneficial type of association with nitrogen-fixing bacteria, aiming to assist in the development of more responsive cultivars to inoculants of beneficial diazotrophs. An integrated differential transcriptome was generated by Illumina RNA sequencing and it provided an overview of sugarcane metabolism, growth and development controlled by nitrogen, water and endophytic nitrogen-fixing bacteria during a successful association. All together, the data suggest that an important control of the efficiency of the association is already set in the early stages of plant-bacterium recognition, when specific plant genotypes sense the environment and regulate several plant signaling pathways involved in microorganism recognition and plant defense.
This work was supported by INCT, CNPq, FAPERJ & CAPES.

W090: Bioenergy Grass Genomics

Field Performance of Sugarcane with TALEN or RNAi Mediated Lignin Reduction

Baskaran Kannan, Je Hyeong Jung and Fredy Altpeter, University of Florida - IFAS, Gainesville, FL

Transcription activator-like effector nuclease (TALEN) is a recently developed tool enabling precise genome modifications, such as targeted mutagenesis, gene replacement, or insertion. Sugarcane is a prime feedstock for bioethanol production, and utilizing both sucrose and cell wall bound sugars for fermentation will enhance the biofuel yield. We recently demonstrated that RNAi mediated downregulation of lignin biosynthetic gene Caffeic acid O-methyltransferase (COMT) is a successful strategy to improve bioethanol production from lignocellulosic sugarcane biomass. In this study, COMT was targeted for the TALEN induced multi-allelic mutagenesis to modify lignin biosynthesis in sugarcane. Targeted mutations following TALEN delivery were identified by capillary electrophoresis of the COMT amplicon. Events were confirmed by sequencing of the COMT amplicon which revealed the presence of insertions and deletions at the target site. Data comparing the total lignin content in the stem biomass of COMT mutant and RNAi suppressed sugarcane events and their field performance will be presented.

W091: Bioinformatics

Structural Variation Analysis with Long Single Molecule Reads

Michael Schatz, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Structural Variants (SVs), which include deletions, insertions, duplications, inversions and chromosomal rearrangements, have been shown to effect organism phenotypes, including changing gene expression, increasing disease risk, and playing an important role in evolution. Still it remains challenging to detect all types of SVs from high throughput sequencing data and it is even harder to detect more complex SVs such as a duplication nested within an inversion.

To overcome these challenges, we have been applying long third generation sequencing reads from Pacific Biosciences and Oxford Nanopore to analyze SVs in a range of samples from small microbial genomes, through mid-sized plant and animal genomes, to large mammalian genomes. The increased read lengths, which currently average over 10kbp and some approach 100kbp, allow us to span more complex SVs and accurately assess SVs in repetitive regions, two of the major limitations when using short reads. To support this analysis, we have developed a new suite of tools for detecting and analyzing SVs from long reads. First, our new read mapping algorithm NGM-LR is able to accurately align the long noisy reads, especially near SVs. Second, our algorithm Sniffles scans the alignments for split-reads and high-mismatch regions to precisely detect SV breakpoints. In complementary work, we use whole-genome alignment to detect certain types of variants directly from a de novo assembly of the long reads. Finally, our algorithm SplitThreader uses a network flow algorithm with the breakpoints and alignment coverage to reassemble the underlying genome structure, even when there are complex nested SVs.

W092: Bioinformatics

Un-zipping Diploid Genome – Revealing the Heterozygous Variants From Comprehensive Haplotig Assembly

Jason Chin[1], Paul Peluso[1], David Rank[1], Maria Nattestad[2], Fritz J. Sedlazeck[2], Michael Schatz[2], Alex Copeland[3], Alicia Clum[3], Kerrie W. Barry[3], Joseph Ecker[4], Ronan O'Malley[4] and Chongyuan Luo[4], (1)Pacific Biosciences, Menlo Park, CA, (2)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (3)DOE Joint Genome Institute, Walnut Creek, CA, (4)Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA

Outside of the simplest cases (haploid, bacteria, or inbreds), genomic information is not carried in a single reference per individual, but rather has higher ploidy (n=>2) for almost all organisms. The existence of two or more highly related sequences within an individual makes it extremely difficult to build high quality, highly contiguous genome assemblies from short DNA fragments. Based on the earlier work on a polyploidy aware assembler, FALCON (https://github.com/PacificBiosciences/FALCON), we developed new algorithms and software ("FALCON-unzip") for *de novo* haplotype reconstructions from SMRT® Sequencing data.

We generate two datasets for developing the algorithms and the prototype software: (1) whole genome sequencing data from a highly repetitive diploid fungal (*Clavicorona pyxidata)* and (2) *whole genome sequencing data from* an F1 hybrid from two inbred *Arabidopsis* strains: Cvi-0 and Col-0.

For the fungal genome, we achieved an N50 of 1.53 Mb (of the 1n assembly contigs) of the ~42 Mb 1n genome and an N50 of the haplotigs (haplotype specific contigs) of 872 kb from a 95X read length N50 ~16 kb dataset. We found that ~ 45% of the genome was highly heterozygous and ~55% of the genome was highly homozygous. We developed methods to assess the base-level accuracy and local haplotype phasing accuracy of the assembly with short-read data from the Illumina® platform. For the *Arabidopsis*F1 hybrid genome, we found that 80% of the genome could be separated into haplotigs. The long range accuracy of phasing haplotigs was evaluated by comparing them to the assemblies from the two inbred parental lines. We show that a more complete view of all haplotypes could provide useful biological insights through improved annotation, characterization of heterozygous variants of all sizes, and resolution of differential allele expression.

The current Falcon-Unzip method will lead to understand how to solve more difficult polyploid genome assembly problems and improve the computational efficiency for large genome assemblies. Based on this work, we can develop a pipeline enabling routinely assemble diploid or polyploid genomes as haplotigs, representing a comprehensive view of the genomes that can be studied with the information at hand.

W093: Bioinformatics

MAKER: The Easy-to-use Genome Annotation Pipeline

M. Yandell, Department of Human Genetics, University of Utah, Salt Lake City, UT

MAKER: The Easy-to-use Genome Annotation Pipeline

Carson Holt, Daniel Ence, Barry Moore, Mark Yandell

University of Utah, Utah Center for Genetic Discover, Salt Lake City UT

MAKER is a portable and easy-to-use genome annotation pipeline and annotation management tool. MAKER identifies repeats; aligns ESTs, RNA-seq data, and proteins to a genome; produces *ab initio* gene predictions, and automatically synthesizes these data into gene annotations with evidence-based quality values. MAKER is easily trainable: outputs of preliminary runs can be used to automatically retrain its gene prediction algorithms, producing higher quality gene-models on subsequent runs. MAKER is also a genome management tool that can update

legacy annotations using new experimental evidence, such as mRNA-seq data; it can also map annotations forward to updated genome assemblies. MAKER's inputs are minimal, and its outputs are GFF3 and FASTA files that can be used by Chado, Gbrowse, JBrowse, or Apollo. We will present highlights from some recently annotated genomes, and explain how we are now integrating MAKER with resequencing efforts to enable rapid genotype-phenotype association.

W094: Bioinformatics

GeneMark Family of Gene Prediction Tools for Prokaryotic and Eukaryotic Genomes

Mark Borodovsky, Shiyuyun Tang and Alexandre Lomsadze, Georgia Institute of Technology, Atlanta, GA

Gene prediction has a fundamental role in genomics. Still, the precise annotation of protein-coding genes in a new genomic sequence is the target difficult to reach.

Prokaryotic genes can be predicted with much higher average accuracy than eukaryotic ones. Nevertheless, the error rate is not negligible and largely species-specific. Most errors are made in prediction of genes located in genomic regions with atypical G+C composition. We present here results of further development of GeneMarkS, a self-training tool used in many genome projects. The new tool, GeneMarkS-2 (Tang et al. in preparation), uses local G+C-specific heuristic models for scoring individual ORFs in the first step of analysis. Predicted atypical genes are retained and serve as 'external' evidence in subsequent self-training iterations. GeneMarkS-2 controls the quality of training process by effectively computing a measure of relative entropy between competing sequence models parameterized by self-training, the measure that guides the selection of optimal model orders.

We also report progress in developing tools for structural annotation of eukaryotic genomes. We are constantly updating the self-training ab initio gene prediction tool, GeneMark-ES. This tool was extended to fully automated GeneMark-ET (Lomsadze et al., 2014) that integrates into the training process mapped RNA-Seq reads as well as to GeneMark-EP (Lomsadze et al. in preparation) that integrates information on footprints of homologous proteins. The automated tool that combines all the options, GeneMark-EX (paper in preparation) selects the mode of running depending on the availability and volume of the types of external evidence (transcripts and proteins).

W095: Bioinformatics

A Widespread, Comparative Assessment of LTR-Retrotransposons - the Major Components of Plant Genomes

Heidrun Gundlach, Helmholtz Center Munich - Plant Genome and Systems Biology PGSB, Neuherberg, Germany and Klaus F.X. Mayer, Helmholtz Center Munich - Plant Genome and Systems Biology, Neuherberg, Germany

By far the major part of plant genomes consists of LTR-retrotransposons and their deteriorated remnants. They can cover over 80% of large genomes, like barley (5 Gbp) or wheat (17 Gbp), adding up to several Gbp of highly repetitive transposon derived sequences. These fast evolving 'junk' sequences define chromosomal architectures, they create phenotypic variations, are involved in epigenetic regulations and can be used to trace back genome evolution. Despite their importance LTR-retrotransposons are not consistently annotated across different plant species. But they are - due to their well defined structural properties - well amenable to bioinformatic de novo detection approaches.

Aimed at efficiently generating homogenous data sets suited for species wide comparisons, we established a streamlined LTR detection and analysis pipeline. The work flow starts with candidate detection via LTRharvest (genometools.org), followed by characterization modules - Pfam domain order for classification, insertion age, family clusters - and finally filter steps to remove non canonical elements. The resulting classified full length LTR-retrotransposon elements are then added to our plant repeat database PGSB REdat (pgsb.helmholtz-muenchen.de/plant/recat/http://pgsb.helmholtz-muenchen.de/plant/recat/) and used as templates to detect the multitude of deteriorated fragments within the genome assemblies.

After a brief introduction to the pipeline principles the talk will present results for a widespread selection of different plant species in respect to basic numbers, insertion age distributions, family sizes and broad to fine scaled chromosomal locations.

W096: Bioinformatics

Complexity of Gene Expression Evolution after Duplication: Protein Dosage Rebalancing

Igor Rogozin, National Institutes of Health, Bethesda, MD

The concept of genetic balance traces back to the early days of genetics. With the increasing availability of genomic data, it became clear that numerous gene families have diverged in function through series of duplications, including many lineage specific expansions (or gene copy-number variations, CNVs, at the population level). This is not surprising taking into account that gene duplications are traditionally considered to be a major evolutionary source of new protein functions Results of large-scale analyses of gene duplications are best consistent with the following possible scenario of gene duplications: many recent gene duplications/CNVs have a positive effect in some tissues and/or environmental conditions, whereas they also have a negative effect in some other tissues and/or environmental conditions. It seems likely that balancing of positive and negative dosage effects is an important factor which is causing diversification of expression patterns (rebalancing of expression) of duplicate genes in the course of fixation of gene duplications.

W097: Brachypodium Genomics

Genome-Wide Alternative Splicing Patterns Modulated During Grass:Virus Interactions

Kranthi K. Mandadi, Texas A&M AgriLife Research, Weslaco, TX and Karen-Beth G. Scholthof, Texas A&M University, College Station, TX

Alternative splicing (AS) influences plant growth, development and response to stress. We analyzed genome-wide AS changes in *Brachypodium distachyon*, infected with *Panicum mosaic virus* (PMV) using RNA-sequencing approach. Approximately 42% of multi-exonic genes in Brachypodium are alternatively-spliced. Among the major AS types, intron-retentions predominated, followed by alternate acceptor, alternative donor and exon-skipping. We identified AS events in ~100 immune-related genes encoding receptor-like kinases, NB-LRR resistance proteins, transcription factors, RNA-silencing and splicing-associated proteins. Cloning and molecular characterization of Bd-*SCL33*, a serine/arginine-rich (SR) splicing factor, uncovered multiple novel intron-retaining splice-variants that are developmentally regulated and modulated during virus infection. Strikingly, Bd-*SCL33* splicing patterns are conserved during infections of additional grass-infecting viruses including *Brome*

mosaic virus, *Barley stripe mosaic virus*, Maize mild mottle virus, Sorghum yellow banding virus, *Wheat streak mosaic virus* and *Foxtail mosaic virus*. Together, these analyses provide new insights into virus-triggered AS landscapes in plants.

W098: Brachypodium Genomics

Modifications of Source-Sink Relationships and the Development of Stress-Tolerant Brachypodium

Nir Sade, UC-Davis, Davis, CA

Environmental stresses are the most serious factors limiting the productivity of agricultural crops. Recent studies have linked the increased frequency of severe abiotic stress events with global warming, emphasizing the urgent need to develop crops with enhanced tolerance to abiotic stresses. Abiotic stress tolerance is particularly important for biomass crops because they must be grown under low-input conditions to maximize the net carbon ratio and on marginal land to minimize competition with food crops. We aim at the development of modules comprising several complementary genes that can be used to engineer stress tolerance into biomass perennial model crops such as *Brachypodium sylvaticum*. Toward this goal, we examined the salt and freezing tolerance of several wild type *B. sylvaticum* lines to determine optimal varieties that could be used as our transgene background. We identified a sensitive (i.e AIN1) and a tolerant varieties (i.e GRE1) for salt and freezing tolerance. GRE1 displayed better biomass and carbon assimilation under high salinity and improved survival under freezing. Preliminary results suggested that sodium compartmentation to the vacuole might be enhanced in GRE1 under stress and this phenomena was well correlated with the expression of a vacuolar NHX1 antiporter. Moreover, vacuolar pH under salt stress was more stable in GRE1 as compared to AIN1. Constitutive over expression of AtNHX1 in the salt sensitive AIN1 background conferred salt tolerance. Quantification of the freezing tolerance of GRE1 and AIN1 will be also presented and discussed.

In the effort to develop broad biotechnological tools for the improvement of perennial grasses, we completed the transformation of AIN1 *B. sylvaticum* with constructs containing 33 genes shown to be associated with enhanced abiotic stress tolerance and source-sink relationships in monocots. Single and/or combination of those genes were overexpressed using either constitutive or stress inducible promoters. We have generated single copy insert homozygous T_2 lines for all constructs. The response of those transgenes lines to various abiotic stresses is being tested.

W099: Brachypodium Genomics

Pan-Genomics in Brachypodium and Implications for Related Grasses

Sean Gordon[1], Bruno Contreras-Moreira[2], David L. Des Marais[3], Diane Burgess[4], Wendy Schackwitz[1], Ludmila Tyler[5], Joel Martin[1], Daniel Woods[6], Anna Lipzen[1], Shengqiang Shu[1], Jeremy L. Phillips[1], Kerrie W. Barry[1], Richard Amasino[6], Ana Caicedo[7], Luis Mur[8], Michael R Freeling[4], Pilar Catalan[9] and John P. Vogel[1], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Fundación ARAID, Zaragoza, Spain, (3)Harvard University, Boston, MA, (4)University of California, Berkeley, CA, (5)Biochemistry and Molecular Biology Department, Amherst, MA, (6)University of Wisconsin-Madison, Madison, WI, (7)Biology Department, University of Massachusetts, Amherst, MA, (8)Aberystwyth University, Aberystwyth,, Wales, (9)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain

Most commonly, intra-species variation is characterized as a set of SNPs and small indels identified by comparing short read data to a single reference genome. In contrast, pan-genomics compares full genome assemblies and gene annotations from multiple individuals of a single species. This approach captures highly divergent and novel sequences not found in the reference genome. Pan-genomics has revolutionized the study of bacterial species, however, few studies have taken this approach with plants, presumably due to the large size and difficulty in accurately assembling plant genomes. To evaluate the feasibility and benefits of plant pan-genomics, we sequenced 54 inbred lines of the model grass, *Brachypodium distachyon*, to a median depth of 92x (10 lines were sequenced with additional 4-6kb mate-pair libraries for scaffolding). Analysis of whole genome de novo assembly and gene annotation reveals that it is possible to recover nearly full genomic sequences from each line, with up to 97% of the whole genome sequence included into syntenic psuedomolecules. Analysis of this data yields a high-confidence *B. distachyon* pan-genome that includes 13,408 core gene clusters found in all lines, 7,283 soft-core genes cluster absent from a few lines, and 17,195 shell gene clusters found in 3 to 52 lines. We find 30,691 gene clusters represented by the reference genome/reference control and a varying number of other genomes. In addition, we identify 7,135 gene clusters not represented in the reference line or controls but present in multiple divergent lines. We show that non-core genes are expressed at lower levels, have narrower and more variable expression across accessions, are evolving faster, have reduced orthology to related grasses and are less likely to have a homeologous gene retained from the ancient genome duplication in the grass lineage. We evaluate the relationship between the number of sequenced lines and their phylogenetic position in relation to the addition of both genic and non-coding sequence. We describe the physical chromosome position of non-core and non-reference genes and its relation to transposable elements. This analysis suggests possible mechanisms by which dispensable genes are eliminated and also barriers to their removal. All lines sequenced in this study are available to researchers.

W100: Brachypodium Genomics

GNRF Represses Floral Transition and Secondary Wall Synthesis in *Brachypodium distachyon*

Pubudu P. Handakumbura[1], Kathryn Brow[2], Sandra P. Romero-Gamboa[2], Scott J. Lee[2], Ian P. Whitney[2], Kangmei Zhao[3], Laura Bartley[3], Henry D. Priest[4], Marion Dalmais[5], Todd C. Mockler[4], Eddy Blondet[5] and Samuel P. Hazen[2], (1)Pacific Northwest National Laboratory, Richland, WA, (2)Biology Department, University of Massachusetts, Amherst, MA, (3)University of Oklahoma, Norman, OK, (4)Donald Danforth Plant Science Center, St. Louis, MO, (5)URGV, Unité de Recherche en Génomique Végétale, Evry, France

Several NAC transcription factors have been shown to play crucial roles in secondary cell wall biosynthesis and overall plant growth and development. *GRASS NAC REPRESSOR OF FLOWERING* (*GNRF*) was selected for functional characterization as it is highly expressed in stems, mirroring an expression pattern similar to that of characterized secondary cell wall regulators such as *Arabidopsis thaliana* NST1, SND1 and SND2. Moreover it is co-regulated with cell wall biosynthesis genes. Over expression mutants (*GNRF-OE*) were generated by constitutively over expressing *GNRF* full-length coding region under the maize ubiquitin promoter. A homozygous mutant allele (*gnrf-1*) harboring a non-synonymous point mutation was isolated from a TILLING mutant collection. Surprisingly, over-expression mutants failed to flower and could

not transition into the reproductive stage. Conversely the *gnrf-1* mutants flowered significantly earlier than control plants. Transcription profiling using a microarray revealed a fifty-fold reduction in two flowering time pathway genes *BdFUL1* and *BdFUL2*, consistent with the persistent vegetative growth phenotype. Three genes involved in cellulose biosynthesis, *BdCESA4/7/8*, two genes associated with lignin biosynthesis, *BdCAD1* and *BdCOMT4* and a gene with a predicted role in xylan biosynthesis *BdGT47-1* were analyzed for changes in transcript abundance. In most cases, these genes were significantly down-regulated in *GNRF-OE* and up-regulated in *gnrf-1* stems. Chemical and histological analysis of stems revealed reduced lignin content compared to the controls where as no significant change was observed between *gnrf-1* and the controls. Taken together, these data suggest that *GNRF* functions as a pleiotropic repressor regulating cell walls and flowering in *Brachypodium distachyon*.

W101: Brachypodium Genomics

How do Rhizobacterial Volatiles Influence Root System Architecture, Biomass Production and Allocation of the Model Grass *Brachypodium distachyon*?

Pierre Delaplace[1], Elena Ormeño Lafuente[2], Minh Luan Nguyen[3], Benjamin Delory[1], Caroline Baudson[1], Magdalena Mendaluk - Saunier de Cazenave[1], Stijn Spaepen[4], Sébastien Varin[1], Yves Brostaux[5] and Patrick du Jardin[1], (1)University of Liège, Gembloux Agro-Bio Tech, Plant Biology Laboratory, Gembloux, Belgium, (2)IMBE (Mediterranean Institut of marine and continental biodiversity) UMR 7263 CNRS, 237 IRD, Marseille, France, (3)University of Liège, Gembloux Agro-Bio Tech, AgricultureIsLife, Gembloux, Belgium, (4)Max Planck Institute for Plant Breeding Research, Department of Plant Microbe Interactions, Köln, Germany, (5)University of Liège, Gembloux Agro-Bio Tech, Statistics, Computer Science and Modeling Laboratory, Gembloux, Belgium

Plant growth-promoting rhizobacteria are increasingly considered as a complement of conventional inputs in agricultural systems. Their effects on their host plants are diverse and include volatile-mediated growth enhancement. The present study aims at assessing the effects of bacterial volatile production on the biomass production and the root system architecture of *Brachypodium distachyon* (L.) Beauv. (line Bd-21). An *in vitro*experimental set-up allowing plant-bacteria interaction through the gaseous phase without any physical contact was used to screen 19 bacterial strains for their growth promotion ability over a 10-day cocultivation period.

Using principal component analysis followed by hierarchical clustering and two-way analysis of variance, five groups of bacteria were defined and characterized based on their combined influence on biomass production and root system architecture. The observed effects range from unchanged to highly increased biomass production coupled with increased root length and branching. Primary root length was only increased by the volatile compounds emitted by *Enterobacter cloacae* JM22 and *Bacillus pumilus* T4. Overall, the most significant results were obtained with *Bacillus subtilis*GB03 which induced a 81% increase in total biomass and enhanced total root length, total secondary root length and total adventitious root length by 88, 196 and 473% respectively.

The analysis of the emission kinetics of bacterial volatile organic compounds is underway and should lead to the identification of volatile compounds candidates responsible for the observed growth promotion effects. Taking into account the inherent characteristics of our *in vitro* system, the next experimental steps are identified and discussed from a fundamental and applied viewpoint.

W102: Brachypodium Genomics

Establishing a Genome-Wide Sequence-Indexed Collection of Brachypodium Mutants

Debbie Laudencia-Chingcuanco[1], Richard Sibout[2], Fabienne Granier[3], Wendy Schackwitz[4], Joel Martin[4], Colin Konishi[1] and John P. Vogel[4], (1)USDA-ARS, Western Regional Research Center, Albany, CA, (2)INRA-IJPB, Versailles, France, (3)INRA, Versailles, France, (4)DOE Joint Genome Institute, Walnut Creek, CA

Brachypodium distachyon has emerged as a powerful model system to address fundamental questions in grass biology. Gaining insights into the function of more than 25,000 genes identified from its sequenced genome will further expand its utility for basic research. To this end, we are establishing a sequence-indexed library of mutations generated by chemical and radiation mutagenesis. Three *Brachypodium* mutagenized populations were produced using ethyl methanesulfonate (EMS), sodium azide (NaN3) and fast-neutron radiation (FNR). With the decreasing cost of DNA sequencing, we utilized whole genome sequencing (WGS) to identify the genomic variations introduced by these mutagens. Results from the pilot sequencing of 91 lines resulted in about 45,000 novel mutations identified. We will discuss the nature of the introduced mutations from chemical and radiation mutagenesis and the lessons we have learned from the process of identifying these new genomic variations.

W103: Brassicas

Homoeologous Recombination Between Duplicated Genes in Brassica Genomes

Xiyin Wang, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

W104: Brassicas

Copy-Number Variation in Flowering-Time Regulators as a Driving Force for Adaptation of *Brassica napus*

Sarah Schiessl[1], Richard Reinhardt[2], Lisa Czaja[2], Diana Kühn[2], Bruno Huettel[2] and Rod Snowdon[1], (1)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (2)Max Planck Genome Centre, Cologne, Germany

The allopolyploid nature of the *Brassica napus* genome enables exchange between different subgenomes, allowing for extensive variation within copy numbers of genes. This is expected to heavily influence traits regulated by affected genes, as demonstrated in numerous animal and plant species. To reveal if this would also prove true for flowering time regulation in *B. napus*, we performed deep sequencing of selected flowering regulators, captured in a diversity panel of 282 genetically divergent accessions, and determined copy number variation (CNV) by assessing coverage depth. Copies from the *B. napus* A subgenome were present in higher numbers than copies from the C subgenome. Some serial CNV events pointed to larger insertions or deletions. We also detected CNV events which seem to be specific for a certain morphotype, for example, a homeoelogous translocation concerning *Bna.FLC* on A10/C09 in swedes (*B. napus* ssp. *napobrassica*), which might be responsible for their

particularly strong vernalisation requirement. The data give novel insight into the consequences of polyploidisation on pan-genomic variation, and will be very useful when considering linkage analysis and genome-wide association studies.

W105: Brassicas

Physiological Genetics of Drought Adaptation in *Brassica napus*

John K. McKay, Colorado State University, Fort Collins, CO and Richard Fletcher, Cargill Specialty Seeds and Oils, Fort Collins, CO

Brassica napus is a globally important oilseed for which little is known about the genetics of drought adaptation. We previously mapped twelve quantitative trait loci (QTL) underlying drought-related traits in a bi-parental cross between winter and spring *B. napus* cultivars. Here we resequence the genomes of the parents to identify genetic diversity across the genome and within QTL regions. We sequenced each parental cultivar on the Illumina HiSeq platform to a minimum depth of 23X and performed reference-based assembly in order to catalog the genomic differences between them. Genome-wide patterns of variation were characterized by higher single nucleotide polymorphism (SNP) density in the A genome and a higher ratio of nonsynonymous to synonymous substitutions in the C genome. QTL analysis with the new, denser map showed the most associated marker to be that developed from an insertion/deletion polymorphism located in the candidate gene *Bna.FLC.A10*, and it was the only candidate within the QTL interval with observed polymorphism. These results provide a glimpse at genome-wide variation differentiating annual and biennial *B. napus* ecotypes as well as a better understanding of the genetic basis of root and drought phenotypes.

W106: Brassicas

Identifying Candidate Genes for Improved Nutrition in Watercress through RNA-Seq

Nikol Voutsina, Adrienne Payne, Mark A. Chapman and Gail Taylor, University of Southampton, Southampton, United Kingdom

Watercress (*Nasturtium officinale* R. Br.; Brassicaceae) is a peppery-flavoured leafy salad crop with a long tradition of cultivation in southern England and increasing popularity around the world. It has received much attention in recent years as one of the most nutrient dense foods and is thought to offer antioxidant and chemopreventive benefits to the consumer. These benefits are derived from the high concentrations of secondary metabolites found in the crop, including glucosinolates. Despite this evidence, watercress remains largely undeveloped with limited breeding resources and no active breeding programmes worldwide. The aim of our work is to understand the genetic basis of important nutritional and agronomic traits in watercress and use modern molecular tools to develop breeding knowledge and resources for the future. Firstly, we will outline the initial steps taken to establish a watercress breeding program at the University of Southampton. Then we will present recent work which utilized RNASeq to assemble and annotate the watercress transcriptome and differential expression analysis to characterize differences between high and low phytonutrient phenotypes. Finally, we will discuss our current efforts to expand molecular tools for this species. Our results, thus far, represent a first and important genetic resource for this understudied crop and include a catalogue of genes known to be, or potentially involved in, nutrient status as well as thousands of markers which could be employed for marker assisted selection and/or genetic mapping for nutritional traits in watercress.

W107: Brassicas

Genomics of Domestication of *Brassica rapa*

Xinshuai Qi, University of Arizona, Tucson, AZ

Genomics of domestication of *Brassica rapa*

Xinshuai Qi[1], Hong An[2], J. Chris Pires[2], Michael S. Barker[1]

1 Dept. of Ecology & Evolutionary Biology, University of Arizona, Tucson, AZ, USA.

2 Division of Biological Sciences, University of Missouri, Columbia, MO, USA.

Understanding the genetic changes associated with domestication is of importance both theoritically and practicially. Genus *Brassica* includes many important crops, like pak choi, Chinese cabbage, turnip, oilseed, broccoli, cauiflower and so on. The vast morphological variations and multiple round whole genome duplications within the genus make *Brassica* crops a good model to study domestication. Pak choi (*B. rapa* var. *chinensis*) and Chinese cabbage (*B. rapa* var. *pekinensis*) are two most important cultivars within *Brassica rapa* (2n = 20), however, their domestication and introduction history are still controversial, and the genetic changes and artificial selection associated with their domestication are still unclear. We sequenced transcriptomes of 162 *Brassica rapa* accessions around the world, with a focus on pak choi and Chinese cabbage in East Asia. Our population genomic analyses and gene differential expression data indicate a single origin of pak choi and Chinese cabbage from China followed with multiple rounds of recent introduction and hybridization events. Gene differential expression patterns in the early developmental stage are consistent with the domestication history and leaf morphological variations. We also identified genes potentially under selective sweep during domestication, which are of importance for breeding and improvement. Further studies will focus on the eastward introduction of *Brassica rapa*from Mediterran, and the influence of hexaploidization on Brassica diversification.

W108: Buffalo genome

Introduction

John Williams, University of Adelaide, School of Animal and Veterinary Science, Roseworthy, Australia

W109: Buffalo genome

Comparative Ruminant Genomics Highlights Segmental Duplication and Mobile Element Insertion Diversity.

Derek Bickhart, Animal Genomics and Improvement Laboratory, ARS-USDA, Beltsville, MD

We have expanded upon a previously reported comparative genomics approach using a read-depth (JaRMs) and a hybrid read-pair, split-read (RAPTR-SV) copy number variation (CNV) detection method that uses read alignments to the cattle reference genome in order to identify species-specific genomic rearrangements. By using a common reference genome, we were able to interrogate alignment discrepancies that were

shared by individuals of different species in order to find putative functional differences such as mobile element insertion (MEI) within transcriptional start sites. Illumina platform sequence reads from four Goat, 14 Water Buffalo and 20 *Bos taurus* Indicus individuals were aligned to the *Bos taurus* UMD3.1 reference assembly. In order to filter misassembled regions and Cattle-specific segmental duplications (SD), we used CNV calls from 40 Bos taurus individuals of 5 different breeds as a background. We were able to confirm previously identified deletions of the Bos taurus bitter taste receptor (T2R65A) and the neuropeptide FF-amid peptide precursor (NPFF) genes in Buffalo with JaRMs and RAPTR-SV results. In just the Buffalo samples, we identified 41 genes that had MEIs within promoter regions after being filtered against our background dataset. The MEI-affected genes showed a DAVID functional enrichment for genes involved in cytoskeletal structure and actin filament organization, suggesting that MEI may be responsible for some of the external phenotypic differences between Buffalo and Cattle. By examining the full spectrum of genomic differences among Cattle, Goat and Buffalo, we will uncover additional species-specific differences that may underlie the unique phenotypes of domesticated ruminant livestock.

W110: Buffalo genome
Extending the Buffalo Annotation
Riccardo Negrini, Associazione Italiana Allevatori, Rome, Italy

W111: Buffalo genome
Diversity of Indian Buffalo
Satish Kumar, Centre for Cellular and Molecular Biology, Hyderabad, India

W112: Buffalo genome
Water Buffalo Genomic Diversity and Post-Domestication Migration Routes
Licia Colli[1,2], Marco Milanesi[1,3], Elia Vajana[1], Daniela Iamartino[4,5], Lorenzo Bomba[1], Francesco Puglisi[6], Marcello Del Corvo[1], Ezequiel Nicolazzi[4], Sahar Saad El-din Ahmed[7], Jesus Rommel V. Herrera[8], Libertado Cruz[9], Shujun Zhang[10], Liguo Yang[10], Xingjie Hao[10], Fuyuan Zuo[11], Song-Jia Lai[12], Shuilian Wang[13], Ruyu Liu[14], Yundeng Gong[15], Mahdi Mokhber[16], Hossein Moradi Shahrbabak[16], Yongjiang Mao[17], Feng Guan[18], Augustin Vlaic[19], Luigi Ramunno[20], Gianfranco Cosenza[20], Ali Ahmad[21], Ihsan Soysal[22], Emel Özkan Ünal[22], Mariena Ketudat-Cairns[23], Jose Fernando Garcia[24], Yuri Tani Utsunomiya[25], Rangsun Parnpai[23], Marcela Gonçalves Drummond[26], Peter Galbusera[27], James Burton[28], Eileen Hoal[29], Yulnawati Yusnizar[30,31], Cece Sumantri[32], Bianca Moioli[33], Alessio Valentini[34], Alessandra Stella[4], John Williams[4,35], International Buffalo Consortium .[4] and Paolo Ajmone Marsan[1,2], (1)Inst. of Zootechnics, Università Cattolica del S. Cuore, Piacenza, Italy, (2)Research Center on Biodiversity and Ancient DNA – BioDNA, Piacenza, Italy, (3)Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ. Estadual Paulista, Araçatuba (SP), Brazil, (4)PTP Science Park, Lodi, Italy, (5)LGS-AIA Associazione Italiana Allevatori, Cremona, Italy, (6)Dipartimento di Scienze Biomediche, Biotecnologiche e Traslazionali, Università degli Studi di Parma, Parma, Italy, (7)National Research Center, Dokki, Giza, Egypt, (8)Philippine Carabao Center, College, Laguna, Philippines, (9)Philippine Carabaou Centre, Nueva Ecija, Philippines, (10)Huazhong Agricultural University, Wuhan, Hubei Province, China, (11)Southwest University, Chang yuan town, China, (12)Institute of Animal Genetics and Breeding, Sichuan Agricultural University, Chengdu, China, (13)Hunan Agricultural University, Changsha, Hunan Province, China, (14)Guizhou Agricultural University, Guiyang, Guizhou Province, China, (15)Enshi Technology college, Enshi, Hubei province, China, (16)University of Tehran, Karaj, Iran, (17)Yangzhou University, Yangzhou, Jiangshu province, China, (18)Hangzhou institute of Metrology of China, Huangzhou, Zhejiang province, China, (19)University of Agricultural Sciences and Veterinary Medicine, Cluj Napoca, Romania, (20)Department of Agriculture, University of Naples “Federico II”, Portici (NA), Italy, (21)COMSATS Institute of Information Technology, Sahiwal, Pakistan, (22)Namik Kemal University, Tekirdag, Turkey, (23)Suranaree University of Technology, Nakhon Ratchasima, Thailand, (24)Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ. Estadual Paulista, Sao Paulo, Brazil, (25)Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ. Estadual Paulista, Jaboticabal, Brazil, (26)Myleus Biotecnologia, Belo Horizonte, Brazil, (27)KMDA, Antwerpen, Belgium, (28)Earthwatch Institute, IUCN, Oxford, United Kingdom, (29)NRF/DST Centre of Excellence for Biomedical TB research, and MRC Centre for TB research, Stellenbosch University, Stellenbosch, South Africa, (30)Research Center for Biotechnology, Indonesian Institute of Sciences, Jalan Raya Bogor, Indonesia, (31)Indonesian Buffalo Conservation and Breeding Center, Ciapus-Bogor, Indonesia, (32)Bogor Agricultural University (IPB), Bogor, Indonesia, (33)Consiglio per la ricerca in agricoltura e l’analisi dell’economia agraria, Monterotondo, Rome, Italy, (34)University of Tuscia, Viterbo, Italy, (35)University of Adelaide, School of Animal and Veterinary Science, Roseworthy, Australia
The 90K Affymetrix Axiom® Buffalo Genotyping Array has been used to genotype river buffalo samples from Pakistan, Iran, Turkey, Egypt, Romania, Bulgaria, Italy, Mozambique, Brazil and Colombia, and swamp buffaloes from China, Thailand, Philippines, Indonesia and Brazil. Model-based clustering algorithms (Admixture and FastStructure software) and graph tools (Treemix and network analysis) have been applied to SNP data to evaluate the levels of molecular diversity and to highlight population structure and migration events. The best-fitting resolution devised by Bayesian clustering highlighted three distinct gene pools in pure river as well as in pure swamp buffalo populations, together with some genomic admixture occurring in the Philippines and in Brazil, in agreement with documented importations of animals for breed improvement purposes. The Mediterranean buffalo and the Carabao breed from Brazil represent the most differentiated gene pools within the river and swamp group, respectively, which is most likely due to genetic bottlenecks, isolation and selection. Gene flow events, evidenced by Treemix and Network analyses, highlighted a likely contribution from the river buffalo gene pool to the admixed swamp populations and, within river buffaloes, from the Mediterranean to the breeds from Colombia and Brazil. When evaluated in a geographical framework, the results of our analyses support archeozoological evidence for the domestication of river and swamp buffalo in the Indian subcontinent and in Southeast Asia, respectively, and furthermore revealed some unexpected patterns of migration, which suggest that the spread of domestic buffaloes out of the domestication center may have followed alternative migration routes.

W113: Buffalo genome

Using the 90K Axiom Buffalo Array for Genome Wide Association Studies

Humberto Tonhati[1], Daniel J. A. Santos[2], Rusbel Raul Aspilcueta-Borquis[3], Gregório M. F. de Camargo[2] and Naudin A. Hurtado-Lugo[2], (1)Unesp, Brazil, São Paulo,, Brazil, (2)Unesp, Brazil, Jaboticabal, Brazil, (3)UFGD, Dourados, Brazil

The buffalo population in Brazil is around 1,332,284 heads. Although it represents just 0.7% of the world population, it is the biggest population in Americas. It has a potential growth since there is a demand for mozzarella production in the country. The aim of the abstract was to sum up the genomic studies that are being developed at the Animal Science Department, Unesp-Jaboticabal-SP, Brazil. Information from 452 buffaloes (*Bubalus bubalis*) (57 sires and 395 dams) genotyped using the 90k Axiom ® Buffalo Genotyping (Affymetrix) and 384 buffaloes (dams) genotyped using the Illumina BovineHD BeadChip (777k) were used in the studies of linkage disequilibrium, genome-wide association and genomic selection. For the Illumina 777k cattle chip, it was possible to verify that ~88% of the SNPs were genotyped in buffaloes, but only ~2% (16,580) of the SNPs were segregating. It means that the majority of them were fixed and don't contribute to the understanding the genetic dynamics of the species. For Affymetrix chip, almost all the SNPs were genotyped and ~66% (61,145) of the SNPs have good quality to be used in genomics selection and association studies. Linkage disequilibrium studies were done using the Affymetrix chip. A r^2 value was 0.30 ± 0.33 (with an average distance of 0.043 ± 0.027 Mb). It means that the SNP information is a suitable tool to obtain genomic breeding values. The LD decreased with increasing physical distance between SNP pairs. The effective size was 86 estimated in the last 10 generations and it is close to the minimum limits recommended by FAO. So, there is a necessity to monitor inbreeding and mating decisions to maintain the genetic diversity. These studies were done using Illumina panel and similar results were found. Genome-wide association studies were done for six production traits and four reproductive traits using the Affymetrix chip. The associations were done with the deregressed breeding values. Major genes were prospect for each trait as well as genes with pleiotropic effect. These studies were also done using the Illumina panel, but different regions were found, mainly because the genome coverage is poorer. Genomic selection studies were done using the Illumina chip. Models with the matrix A and a model with the matrix H were compared for six production traits. The heritabilities estimates were similar, but the accuracies of the breeding values were higher for the models with the matrix H. The lack of a buffalo genome makes the genetic progress and understanding of the species a bit slower. The idea of the group is to genotype more animals in order to increase the accuracy of the estimates predicted and repeat the analyses done as well as develop others. We look for collaborators who may contribute with the current buffalo genomic project.

W114: Buffalo genome

Identification of the Causal Mutation for Transverse Hemimelia in Mediterranean Italian River Buffalo Using Whole Genome Sequence Data

Lynsey K. Whitacre[1], Jesse L. Hoff[2], Jared E. Decker[2], Robert D. Schnabel[2], Sara Albarella[3], Francesco Strozzi[4], Chiara Ferrandi[4], Francesca Ciotola[5], Vincenzo Peretti[5], John Williams[6] and Jerry Taylor[7], (1)Informatics Institute, University of Missouri, Columbia, MO, (2)Division of Animal Sciences, University of Missouri, Columbia, MO, (3)Department of Veterinary Medicine and Animal Production, University Federico II - Naples - Italy, Naples, Italy, (4)Parco Tecnologico Padano, Lodi, Italy, (5)Department of Veterinary Medicine and Animal Production, University Federico II, Naples, Italy, (6)University of Adelaide, School of Animal and Veterinary Science, Roseworthy, Australia, (7)University of Missouri, Columbia, MO

Mediterranean river buffalo have recently undergone strong selection for increased quality and quantity of milk for mozzarella cheese production. Strong selection for traits such as milk production are often associated with increased inbreeding, leading to decreased genetic diversity and an increase in the prevalence of genetic diseases. Transverse hemimelia (TH) is a congenital developmental abnormality characterized by the absence of a variable portion of the distal limbs. It occurs at the rate of approximately 2-5% in Mediterranean river buffalo populations and causes significant production loss in affected animals as well as in carriers of the disease, which are eliminated from the breeding herd once identified. Thus, there is significant motivation to identify the causal mutation for TH and develop a genetic test to identify carriers and implement a selective breeding program to eliminate carrier-to-carrier matings. In order to localize the mutation, the genomes of 4 cases and 14 carriers were sequenced using Illumina technology with paired-end libraries and an average depth of 10X coverage. Variant calling from whole genome sequence data resulted in 17.9 million high confidence single nucleotide polymorphisms (SNPs) to be analyzed. Because TH is believed to be an autosomal recessive disorder, variants were only selected for further analysis if all cases were homozygous for the non-reference allele at the SNP and if the frequency of this allele was <30% in the controls. This drastically reduced the number of candidate variants to 459 SNPs. Annotation of these variants identified a single candidate gene, *SMARCA4*, which had previously been implicated in embryonic limb development. Further investigation of *SMARCA4* revealed that it is contained within a run of homozygosity in the affected animals. Ongoing research, including *de novo* assembly for structural variant detection and variant effect prediction, is underway to elucidate the causal mutation.

W115: Buffalo genome

Close

John Williams, University of Adelaide, School of Animal and Veterinary Science, Roseworthy, Australia

W116: Cacao Genomics Workshop

Translating Targeted Mutagenesis Technologies into Crop Species

Shaun J Curtin[1], Nevin D. Young[1], Dan Voytas[2] and Robert M. Stupar[3], (1)Department of Plant Pathology, University of Minnesota, St. Paul, MN, (2)University of Minnesota, Saint Paul, MN, (3)Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

Several methodologies have recently been developed to target specific DNA modifications within a genome. Researchers in a wide range of crop species are developing knockout or gene deletions at specific loci for desirable trait outcomes. This talk will address recent advances in developing targeted mutagenesis systems in soybean and Medicago. Soybean has seven functional dicer-like (Dcl) genes as a result of past

duplication events, including a single-copy Dcl3 locus and two copies each of Dcl1, Dcl2 and Dcl4. These genes function in RNA processing and are critical for proper RNA silencing and regulation of gene expression. To study their specific (and potentially redundant) roles in the RNA silencing pathways, we have targeted each soybean Dcl gene with either a zinc-finger nuclease (ZFN), a TAL-effector nuclease (TALEN) or a CRIPSR/Cas9 nuclease designed to induce mutations. To date, we have generated non-transgenic lines harboring mutations for five of the seven Dcl genes. In addition, we have used the same platform to engineer Medicago mutants to validate candidate genes associated with symbiotic nitrogen fixation (SNF) previously identified by genome wide association studies (GWAS). This work demonstrates the capacity to develop novel mutant alleles though targeted mutagenesis and derive new phenotypes through stacking these alleles. These efforts have also provided insight into the opportunities and potential obstacles that exist in the translation of gene targeting tools into crop species.

W117: Cacao Genomics Workshop

Improving Fruit and Vegetable Flavors with 'Consumer Assisted Selection'

Kevin M. Folta, University of Florida, Gainesville, FL

Matching fruit and vegetable sensory attributes to human sensory desires is a difficult task. Plant product quality is shaped by interactions between genetics, environment and management. Human preferences are formed from interaction between sensory receptors, personal experiences, and marketing-based perceptions. The overlap between these two areas defines the window of consumer liking for plant products. Traditional plant genetic improvement has sought to increase production qualities such as disease resistance, yield, shipping quality and shelf life. The remarkable success as breeding for these production traits, while de-prioritizing sensory quality, has led to a produce market offering room for flavor and aroma improvement. To meet this challenge the Plant Innovation Center at the University of Florida brought together non-traditional collaborators. Research groups from human sensory analysis, statistical scaling, genomics, biochemistry, marketing, postharvest physiology, plant breeding and other disciplines coalesced around the question of defining consumer preferences, and then using technology to breed new plant varieties to meet those objectives. The results are new cultivars specifically designed to the demands of the consumer, which satisfy production requirements. This 'consumer-assisted selection' approach requires integrated expertise working in a coordinated network, as the nexus of plant and human complexities cannot be addressed one dimensionally.

W118: Cacao Genomics Workshop

CRISPR Genome Editing in Outcrossing Woody Perennials: Living with SNPs

Chung-Jui Tsai, University of Georgia, Athens, GA

The CRISPR/Cas9 technology is revolutionizing all facets of biology from medicine to agriculture, owing to its precision, efficiency, versatility and ease of adoption. We recently reported the first application of CRISPR/Cas9 in stably transformed *Populus*, extending the species range of this powerful technology to woody perennials. We achieved 100% mutational efficiency in two 4-coumarate:CoA ligase (4CL) genes tested—all stably transformed plants contain bi-allelic DNA modifications. In both cases, the primary transformants exhibited strikingly uniform phenotypes among independent transgenic lines, a consistency that is unmatched by previous gene silencing methods. An underappreciated obstacle in genome editing of outcrossing species is the frequent occurrence of sequence polymorphisms that can render CRISPR/Cas9 unproductive. However, current genome data mining tools and resources do not support exploration of sequence variations. We present experimental evidence as well as genome-wide computational analysis to demonstrate the sensitivity of CRISPR/Cas9 to allelic heterozygosity, and discuss tools and strategies that can help deal with such sequence polymorphisms. With its specificity, CRISPR/Cas9 offers a less equivocal means than previous approaches for discerning functional redundancy of paralogous genes that are prevalent in plant genomes. Continuing improvements of the CRISPR/Cas9 system for multiplex genome engineering should facilitate these efforts. For woody perennials such as forest trees, fruit/nut trees and woody ornamentals with long generation times, CRISPR/Cas9 affords a facile means for precision gene editing to accelerate not only basic research but also crop improvement.

W119: Cacao Genomics Workshop

Differential Cacao Cultivar Gene Expression with RoDEO

Niina Haiminen[1], Laxmi Parida[1] and David N. Kuhn[2], (1)IBM T J Watson Research - Computational Biology Center, Yorktown Heights, NY, (2)USDA ARS SHRS, Miami, FL

Moniliophthora roreri, a fungal pathogen of cacao, causes frosty pod (FP) disease where fungal mycelial growth on the pod appears similar to frost. To identify genes that could be critical to resistance to FP, we performed a time course of inoculation of a susceptible cultivar (Pound 7) and a resistant cultivar (UF273) with FP conidia.

Pods on the trees were inoculated or mock inoculated, harvested at 8, 24 and 48 hours, RNA was isolated and sequenced on an Illumina HiSeq. The gene expression data were analyzed with RoDEO (Robust Differential Expression Operator), see http://researcher.watson.ibm.com/project/5513. RoDEO employs a non-parametric method assessing stable as well as differentially expressed genes, and is particularly resilient when dealing with a limited number of replicates.

We observed robust cultivar-specific expression patterns that persist across time points and individuals plants. In addition, we identified genes whose expression is consistently different between the two cultivars, including genes linked to disease resistance.

W120: Cacao Genomics Workshop

Evaluating the impact of mislabeled accessions in a cacao progeny trial

Ashley Duval, Mars, Inc., Miami, FL

The issue of offtypes (misidentified accessions) in germplasm collections of cacao (*Theobroma cacao* L.), presents a significant challenge to breeders in efforts to improve generational gains in key traits like production and disease resistance. It has been estimated that anywhere from 13-40% of global germplasm collections are contaminated with offtypes. Breeding programs rely on the assumption that parents have been correctly characterized and identified to develop the best families and select the best progenies. Molecular fingerprints from a progeny trial (PT08) at the Mars Center for Cocoa Science in Ilhéus, Brazil were used to verify the identities of 3,436 progeny from 69 crosses made according to a Circulant Factorial design and planted in 2009. Three years of phenotypic data including yield, pod index, seed index, percent of

healthy pods, and Witches Broom Disease (*Moniliopthora perniciosa*) was analyzed using linear mixed modeling in ASREML. Analyses were made with and without the exclusion of offtypes to estimate their effect on selections, prediction of genetic gains, and the inferred genetic architecture of phenotypic traits of interest. Different approaches to handling offtypes through data analysis are compared, such as use of a pedigree matrix. Additional approaches to improving reliability and accuracy of trait estimations through modifications to the trial designs and analyses are discussed.

W121: Cacao Genomics Workshop

Bioinformatics of cacao

Joseph C. Stack, Mars, Inc., Miami, FL

W122: Camelids

Genomic Footprints of Selection under Domestication in Old World Camelids

Robert Fitak, Institute of Population Genetics, Vetmeduni Vienna, Austria and Department of Biology, Duke University, Durham, NC, Durham, NC, Elmira Mohandesan, Vetmeduni Vienna, Institute of Population Genetics, Vienna, Austria, Jukka Corander, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland, Adiya Yadamsuren, Mongolian Academy of Sciences, Institute of Biology, Mammalian Ecoloy Laboratory, Ulaanbaatar, Mongolia, Battsetseg Chuluunbat, Laboratory of Genetics, Institute of Biology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia, Omer Abdelhadi, University of Khartoum, Department for Meat Sciences, Khartoum, Sudan, Abdul Raziq, Lasbela University of Agriculture, Water and Marine Sciences, Uthal, Pakistan, Bernard Fayes, CIRAD-ES, Montpellier, France and **Pamela A. Burger**, Research Institute of Wildlife Ecology, Vetmeduni Vienna, Austria, Vienna, Austria

In many parts of the Old World, domesticated camels (genus *Camelus*) are an essential resource, providing food, labor, commodities, and sport to millions of people. Of the three extant species, two have been domesticated (single-humped dromedaries, *Camelus dromedarius*, and two-humped Bactrian camels *Camelus bactrianus*) and one remains wild (two-humped wild Bactrian camels *Camelus ferus*). All three species possess a variety of adaptations to harsh desert conditions, including mechanisms to tolerate extreme temperatures, dehydration, and sandy terrain. Recent genomic studies of camels have identified patterns of selection consistent with the aforementioned adaptations in addition to quantifying genetic variation and examining demographic history. However, these studies are limited to analyses based upon a single genome from each species, thus biasing many inferences of selection and adaptation. Furthermore, draft genomes are extremely susceptible to errors in the estimated number of genes, thereby distorting conclusions of adaptation based upon orthologous genes between species. In this study, we take a population genomics approach to inferring both positive selection and demographic history of Old World camelids. By re-sequencing multiple genomes from all three species, our objectives were to i) identify genes or regions under selection within and between species related to domestication and/or adaptation, ii) examine the recent demographic history and genome ancestry, and iii) provide an extensive set of genomic resources for future studies of camels.

W123: Camelids

Characterisation of the Genetic Diversity, Structure and Admixture of Dromedary Populations

Faisal Almathen, King Faisal University, Al-Hasa, Saudi Arabia, Pauline Charruau, Institute of Population Genetics, Vienna, Austria, Joram M Mwacharo, International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia, Pamela A. Burger, Research Institute of Wildlife Ecology, Vetmeduni Vienna, Austria, Vienna, Austria and Olivier Hanotte, The University of Nottingham, Nottingham, United Kingdom

The dromedary (*Camelus dromedarius*) commonly referred to as 'ship of the desert' has played an important part in the development and expansion of trading networks across inhospitable habitats over three millennia. Caravan roads were among the main exchange networks in human history and they likely facilitate livestock movements across large geographic distances. In particular dromedary camels are known to have been traded along these roads. The aim of this study was to use genetic markers to assess the population genetic diversity and structure of the *Camelus dromedarius*. Nine hundred and seventy animals from 20 countries representative the entire species geographic range were screened with 17 microsatellite loci. Our results support extensive dromedary camel movements along the trade routes contributing significantly to the current gene pools. More particular, weak genetic differentiation is observed across most of dromedary populations, likely the result of high gene flow along the trading routes. However, the Horn of Africa population appears to have been genetically isolated from the others populations, a possible consequence of its environmental and cultural distinctiveness limiting the amount of gene flow. The results foster our comprehension of the evolutionary history of *C. dromedarius* while providing an unique new window on the pattern and process of trading in ancient times.

W124: Camelids

Preliminary Population Structure Analysis Using Whole Genome and Genotyping-by-Sequencing Data of Dromedary Camels

Heather M. Holl[1], Salma Abdalla[2], Benjamin M. Shykind[2], Joel A. Malek[2], Yasmin A. Mohamoud[3], Donald Miller[4], Kamaal Pasha[5], Adel Khalili[5], Douglas Antczak[4] and Samantha A. Brooks[1], (1)University of Florida, Gainesville, FL, (2)Weill Cornell Medical College, Doha, Qatar, (3)WCMC-Qatar, Doha, Qatar, (4)Cornell University, Ithaca, NY, (5)Tharb Camel Hospital, Doha, Qatar

The domestic dromedary camel is of economic and cultural importance in many countries in north Africa, the Middle East, and in parts of Asia. With a global population estimate of over fifteen million, many in developing nations, camels are often selected for meat and milk production, draught, riding, and racing traits. Camels possess an array of unique physiological adaptations that have enabled their survival in the arid desert environment. Assembled using only Illumina reads, the currently available reference genome contains 32,572 scaffolds with a N50 of 4.2Mb. We plan to produce a more robust set of genomic resources that will facilitate genomic selection and the study of desert adaptation in mammals using a comparative genomic approach. Genetic diversity and population dynamics will be assessed by genotyping camels from unique subpopulations.

In a preliminary sample set, we collected whole blood samples from dromedary camels located in the US and Qatar. We performed whole genome sequencing on nine camels (four US and five Qatari) and genotyping-by-sequencing (GBS) on twenty-four Qatari racing camels. Preliminary analysis of the GBS reads identified 310,311 SNPs prior to filtering. Variants will be called separately on both data sets, then filtered to those detected by both technologies for further analysis. Here we present genetic diversity metrics assessed with standard population genetic pipelines. Analysis of the sequencing data is currently ongoing.

W125: Camelids

Seasonal Adaptations of the Hypothalamo-Neurohypophyseal System of the Arabian One-Humped Camel

David Murphy[1], F. Zahra Djazouli Alim[2], Charles Hindmarch[1], Michael Greenwood[1], Mark Rogers[1], Chan Kok Gan[3], Tay Yealing[4], Elena V. Romanova[5], Bruce R. Southey[5] and Jonathan V Sweedler[5], (1)University of Bristol, Bristol, England, (2)Université Saad Dahleb, Blida, Algeria, (3)University of Malaya, Kuala Lumpur, Malaysia, (4)BioEasy Sdn Bhd, Shah Alam, Selangor Darul Ehsan, Malaysia, (5)University of Illinois at Urbana-Champaign, Urbana, IL

The “ship of the desert”, the one-humped Arabian camel (*Camelus dromedarius*), has a remarkable capacity to survive in conditions of extreme heat without needing to drink water. One of the ways that this is achieved is through the actions of the antidiuretic hormone vasopressin (AVP) and the natriuretic hormone oxytocin (OXT), both of which are made in a specialised part of the brain called the hypothalamo-neurohypophyseal system (HNS), but exert their effects at the level of the kidney to, respectively, provoke water conservation and salt excretion. Interestingly, our electron microscopy studies have shown that the ultrastructure of the camel HNS changes according to season, suggesting that in the arid conditions of summer the dromedary’s HNS is in a state of permanent activation, in preparation for the likely prospect of water deprivation. We have sequenced the genome of an Algerian camel called “Jamal”; his genome consists of 2.09Mb (210,368 contigs, with 115,325 >1000nt) from which we predicted 64,989 gene models. Based on this unique resource, we have carried out a preliminary RNAseq analysis of the camel HNS in summer and winter. Interestingly, amongst the 435 genes found to be significantly differentially regulated (edgeR p value <0.05) is OXT (winter 4.3 fold greater than summer). Further, we have shown that the AVP gene is subject to unique camel specific alternative splicing events that may result in the production of novel peptides. Ongoing peptidomic studies are relating these findings to the elaboration of hormones by the camel HNS in response to season.

W126: Camelids

Improving the Alpaca Genome Sequence Assembly

Terje Raudsepp[1], Mark Richardson[2], Fahad A Alshanbari[3], Polina Perelman[4] and Belinda Appleton[2], (1)Texas A&M University, College Station, TX, (2)Deakin University, Victoria, Australia, (3)TAMU, College Station, TX, (4)Institute of Molecular and Cellular Biology, Novosibirsk, Russian Federation

The development of whole genome sequence maps is key for understanding the genome architecture of a species and is critical for the discovery of genetic blueprints of diseases, congenital disorders and traits of biological significance. Here we report our recent work improving the alpaca VicPac2.0 reference assembly by incorporating 100X Illumina HiSeq data from the same female individual. Annotation of the assembly was improved by mapping alpaca testis RNA-Seq reads to sequence scaffolds. This allowed discovery of missing exons, determining new splice sites and gene isoforms. RNA-Seq reads that did not map to the female alpaca genome sequence scaffolds were assembled *de novo,* resulting in 210,961 transcripts representing novel autosomal and X-linked genes, and the first 44 putative Y chromosome transcripts. The improved sequence scaffolds are currently anchored to alpaca chromosomes using the recently generated whole genome cytogenetic map and by additional FISH mapping of unassigned scaffolds. Our ultimate goal is to create a robust, annotated and chromosomally assigned sequence assembly for the alpaca genome.

W127: Camelids

The Expression Level of the Alpaca *Agouti* Gene has a Marked Impact on Fibre Colour Intensity

Kylie Munyard, Curtin University, Perth, Australia

Most of the research into the molecular aspects of alpaca colour genetics has focussed on the interactions between the melanocortin-1 receptor and its ligand agouti. This information has improved our understanding of how the switch between eumelanin and pheomelanin production is controlled in alpacas. However it does not explain all of the many of different colours observed in the species, nor is there an understanding of what controls intensity of colour.

We used RNAseq to determine the expression levels of genes expressed in the skin of white, light-skinned fawn, dark-skinned fawn, light bay, bay and black alpacas in order to identify genes that may be important in controlling fibre colour intensity in the species. These data show that *agouti* may play a role in controlling pigment intensity, as well as its role in pigment switching. *Agouti* is most highly expressed in white alpacas (RPKM=120), is moderately expressed in fawn and bay phenotypes (RPKM=44-69), and has low expression in the black phenotype (RPKM=4). Six different alternate non-coding exon 1 *agouti* sequences were identified. None of the alternate agouti non-coding exon 1 sequences are restricted to any particular colour or group of colours. Interestingly, the two most common exon 1 isoforms map to the genome approximately 400Kb upstream of the *agouti* gene, in the incorrect orientation for a long distance alternate splicing event. Expression of *tyrp1*, a key gene in the production of eumelanin, is negatively correlated with *agouti* expression.

W128: Cassava Genomics

Insights from Genome Sequencing of Cassava and other *Manihot*

Jessica B. Lyons, University of California, Berkeley, Berkeley, CA

We characterized cassava genetic diversity via whole genome shotgun sequencing of 58 accessions, including a global collection of cassava, wild relatives (*Manihot esculenta* ssp. *flabellifolia*), and other *Manihot*. We confirm and date the whole-genome duplication that occurred prior to the split between *Manihot* and *Hevea*, and date the *M. esculenta—M. glaziovii* divergence. We find widespread interspecific admixture that raises questions about the nature of *Manihot* species, and we show that an *M. glaziovii*haplotype is shared between cassava accessions whose ancestry

can be traced back to the Amani program of the 1930's. By examining a panel of 268 African cassava accessions we reveal that this haplotype is enriched among improved cassava varieties.
Progress on cassava genome assembly version 7, based on PacBio sequencing; and draft assemblies for *M. esculenta* ssp. *flabellifolia* and *M. glaziovii*, will be presented. Large-scale structural genomic variation between cassava, *M. esculenta* ssp. *flabellifolia*, and *M. glaziovii* will be explored.

W129: Cassava Genomics
CG Gene Body DNA Methylation Changes and Evolution of Duplicated Genes in Cassava
Steve Jacobsen, University of California at Los Angeles, Los Angeles, CA
DNA methylation is important for the regulation of gene expression and the silencing of transposons in plants. Here we present genome-wide methylation patterns at single-base pair resolution for cassava (*Manihot esculenta*, cultivar TME 7), a crop with a substantial impact in the agriculture of subtropical and tropical regions. On average, DNA methylation levels were higher in all three DNA sequence contexts (CG, CHG, and CHH, where H equals A, T or C) than those of the most well studied model plant *Arabidopsis thaliana*. As in other plants, DNA methylation was found both on transposons and in the transcribed regions (bodies) of many genes. Consistent with these patterns, at least one cassava gene copy of all of the known components of Arabidopsis DNA methylation pathways was identified. Methylation of LTR transposons (*GYPSY* and *COPIA*) was found to be unusually high compared to other types of transposons, suggesting that the control of the activity of these two types of transposons may be especially important. Analysis of duplicated gene pairs resulting from whole genome duplication showed that gene body DNA methylation and gene expression levels have co-evolved over short evolutionary time scales, reinforcing the positive relationship between gene body methylation and high levels of gene expression. Duplicated genes with the most divergent gene body methylation and expression patterns were found to have distinct biological functions and may have been under natural or human selection for cassava traits.

W130: Cassava Genomics
Variant Prioritization for Genomic Selection in Cassava
Dunia Pino del Carpio, Cornell University, Ithaca, NY
With the advent of next generation sequencing Genomic prediction and Genome wide association studies are the leading approaches in the use of whole genome marker and phenotypic data. Results coming from these approaches have been considered independently from each other and used aiming at different goals.Through whole-genomic prediction genotypic or breeding values for both observed and unobserved individuals can be predicted. While GWAS studies aim to dissect the genetic architecture detecting markers significantly associated to traits with a defined threshold.
To date, prediction models in plants have used *all* markers across the genome for prediction assuming *equal* contribution and tend to avoid marker selection. In this scenario in Cassava moderate to low prediction accuracies values have been observed
In previous studies a generalized way to building trait-specific genomic relationships matrices, using GWAS results via BLUP, has been tested using animal and plant models.Additionally, statistical models can be improved if structural genome annotation is included with subsets of annotated SNPs as GRM-kernels.
Within the scope of the present study we included genome-wide association results (p-values and marker effects) of root weight, root number and as dry matter as weights for the single nucleotide polymorphism (SNP) markers when constructing the genomic relationship matrix (GRM) for genomic prediction. Moreover we followed a multi-kernel approach partitioning SNPs based on their statistical significance/or not in association mapping studies and biological/functional annotation.
Results and implications of the study in terms of genomic prediction accuracies are discussed .

W131: Cassava Genomics
Automated Phenotyping for Disease Measurement in Cassava Plants using Smartphones
Ernest Mwebaze, Makerere University, Kampala, Uganda
Cassava is an important security crop in most of Africa. Crop diseases are one of the leading causes of low yields in cassava particularly amongst small-holder farmers. Disease identification is commonly done by experts. In this work I discuss some methods we are using to speed up diagnosis and measurement of disease by automating the phenotyping of cassava crops. We do this using computer vision techniques implemented on smartphones. I will discuss the measurement of disease from symptoms that are manifest on the leaves and the root cross-sections of cassava plants.

W132: Cassava Genomics
Assessing Diversity in Cassava through the Application of Metabolomics
Paul D. Fraser, Royal Holloway University of London, Egham, United Kingdom

W133: Cassava Genomics
Unlocking the Breeding Potential of African Crops through Efficient Data Management: The Example of Cassavabase
Guillaume Jean Bauchet[1], Isaak Y. Tecle[1], Naama Menda[1], Alex C. Ogbonna[1,2], Bryan Ellerbrock[1], Nicolas Morales[1], Agbona Afolabi[3], Ismail Rabbi[3], Peter Kulakow[3], Jean-Luc Jannink[4], Jeremy D. Edwards[5] and Lukas Mueller[1], (1)Boyce Thompson Institute for Plant Research, Ithaca, NY, (2)National Root Crops Research Institute (NRCRI), Umuahia, NY, Nigeria, (3)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (4)Cornell University, Ithaca, NY, (5)USDA-ARS, Stuttgart, AR
Cassava (*Manihot esculenta*) is the main source of calories for 250 million people in Africa and therefore indispensable for food security. Cassavabase (www.cassavabase.org) is the central datastore for the NextGen Cassava Project, aiming to leverage cassava genomic breeding and accelerate variety development in Africa (www.nextgencassava.org/). Cassavabase aims to provide a complete bioinformatic toolset to the

cassava research community. From the field to the lab to the office, it offers integrative solutions for information tracking, streamline management of genotypic and phenotypic data and genomic selection analyses.
Through its web interface, Cassavabase is a globally accessible resource, enhancing data sharing and communication within cassava research community in Africa and overseas. User feedback drives new development and improvements through an active communication system. Current tools cover phenotyping data (trait ontology dictionary, fieldbook app), sequencing data (GBS pipeline, Jbrowse), breeding management (breeding program, trial, trait, pedigree search) and molecular breeding (trait descriptive statistics, genetic maps, genomic selection).
Since its inception, 7 million phenotypic datapoints and 18,000 genotyped accessions have been stored from 500+ users across half dozen breeding programs. By the project end, Cassavabase will be fully hosted at IITA, Ibadan, Nigeria, providing a "all in one" solution for cassava researchers and breeders worldwide. All current developments and data are open source and available to the community (https://github.com/solgenomics/).
In this presentation, we will provide an overview of the current data and tools in Cassavabase. We will highlight challenges encountered and future developments.

W134: Cat & Dog Workshop
99 Lives Cat Whole Genome Sequencing Initiative Update, nearly there!
Leslie A. Lyons, University of Missouri-Columbia, Columbia, MO
The 99 Lives cat whole genome sequencing initiative is a research community-based project that will sequence the genomes of cats to: 1) Identify normal and abnormal genetic variation, 2) Identify causative mutations for specific health concerns, 3) Allow veterinary hospitals to provide individual genome sequencing for cats for state-of-the-art health care. Illumina HiSeq 100 bp paired end sequencing reads from two PCR-free libraries per cat of 350 bp and 550 bp for 20X – 30X genome coverage is suggested. Maverix Biomics provides the bioinformatics support, aligning the reads to the cat genome Felis_catus-6.2 sequence and using Platypus to call variants. The cat reads and SNPs are overlaid onto a UCSC-type browser for easy viewing of the data. Data tables with specific filters can be accessed that provide the identified SNPs in specific genes, regions, chromosomes, or the entire genome. Once a genome dataset is transferred to Maverix, the contributor will be given access to the website for the *99 Lives* cat genome data. Contributors are expected to provide the basic signalment of the cats when possible, such as gender, breed, place of origin, and coat color. All cats are welcome, including exotic and wild felids. The current dataset has 51 domestic cats and 4 wild felids. The new analysis will include 82 felids, comprised of 11 exotic felids and 71 domestic cats. The most recent complete database analysis will be performed in late fall 2015. Additional cat sequencing will likely employ updated illumina-based technologies.

W135: Cat & Dog Workshop
New Resources for Canine GWAS
Adam R. Boyko[1], Jessica Hayward[1], Laura Shannon[1] and Pavel Korniviel[2], (1)Department of Biomedical Sciences Cornell University, Ithaca, NY, (2)Cornell University, Ithaca, NY
Updated Illumina CanineHD array to improve coverage of genome caps and coding regions.

W136: Cat & Dog Workshop
Pursuing the Origins of Dog Domestication using Palaeopopulation Genomics
Greger Larson, University of Oxford, Durham, United Kingdom
Dogs were unquestionably the first domestic animal and the only animal domesticated within a hunter-gatherer context prior to the advent of agriculture. Understanding the precise temporal and geographic origins of domestic dogs has proven difficult for several reasons including: the widespread distribution of wolves and the lack of an easily interpretable phylogeographic signatures amongst modern dog populations. More recently, studies making use of high-coverage genomes of dogs and wolves have demonstrated that the wolf population from which all dogs descend is likely extinct, exacerbating the difficulty in identifying the wolves which gave rise to dogs. In addition, the history of both domestic plants and animals has incorporate significant degrees of admixture between domestic animals and wild populations that were never involved in the original domestication process. Here I present an empirical demonstration of long-term admixture and patterns of mitochondrial turnover within ancient dog samples. In addition, I will describe a preliminary analysis of a >25x nuclear genome of a Neolithic European dog and discuss how this sample demonstrates the power of ancient genomic data from global samples of ancient dogs and wolves to piece together where, when and how many times dogs were domesticated.

W137: Cat & Dog Workshop
Explore the Truth About Cats and Dogs with Affymetrix
Mohini Patil, Affymetrix, Santa Clara, CA
Genotyping solutions from Affymetrix provide breeders and researchers with powerful and flexible tools to identify, screen and validate complex genetic traits in animals. These include both arrays and next generation sequencing (NGS)-based assays. Axiom® Genotyping Solution is fully automated and can process more than eight sample plates per week with a single instrument. It allows for high throughput screening of markers of interest in 96-array format, with selection of markers of interest in 384-array format. This platform can be used to for phenotype-trait association and selection applications. At the other end of the spectrum, many researchers need an affordable high throughput method to provide customizable genotyping for interrogation of tens to thousands of markers per sample. Eureka™ Genotyping Solution utilizes NGS to enable genotyping of thousands of DNA samples for tens to thousands of loci. The Eureka genotyping assay is a ligation-dependent PCR reaction, which uses interrogation site bar codes contained within the ligation probes, as well as sample index bar codes added during the amplification step. Next generation sequencing libraries can be created for thousands of DNA samples within 24 hours. The Eureka genotyping assay has been used to interrogate simple and poly allelic SNPs, small to large INDELs, tetra-ploids and for the detection of low percent contamination in a large background of non-target DNA. The Eureka Genotyping Solution is flexible, accurate, specific, affordable and robust.

W138: Cat & Dog Workshop

Evidence of Selection Signatures that Shape the Persian Breed

Francesca Bertolini[1], Barbara Gandolfi[2], Eui-Soo Kim[1], Bianca Haase[3], Leslie A. Lyons[2] and Max F. Rothschild[1], (1)Iowa State University, Ames, IA, (2)University of Missouri-Columbia, Columbia, MO, (3)University of Sydney, Sydney, Australia

The Persian cat is mainly characterized by an extremely brachycephalic face as part of the standard body conformation. Despite the popularity, world-wide distribution and economic importance of the Persian cat as a fancy breed, little is known about the genetics of their hallmark morphology, brachycephaly. Over 800 cats from different breeds including Persian, non-Persian breeds (Abyssinian, Cornish Rex, Bengal, La Perm, Norwegian Forest, Maine Coon, Manx, Siamese, and Oriental) and Persian-derived breeds (British Shorthair, Scottish Fold, Selkirk Rex) were genotyped with the Illumina 63K feline DNA array. The experimental strategy was composed of three main steps: i) the Persian dataset was screened for runs of homozygosity to find and select highly homozygotes regions; ii) selected Persian homozygous regions were evaluated for the difference of homozygosity between Persians and those considered non-Persian breeds, and, iii) the Persian homozygous regions most divergent from the non-Persian breeds were investigated by haplotype analysis in the Persian derived breeds. Four regions with high homozygosity (H > 0.7) were detected, each with an average length of 1 Mb. Three regions can be considered unique to the Persian breed, with a less conservative haplotype pattern in the Persian-derived breeds. Moreover, two genes, *CHL1* and *CNTN6* known to determine face shape modification in humans, reside in one of the identified regions and therefore are candidates for the brachiocephalic face in Persians. In total the homozygous regions contained several neuronal genes that can be involved in the Persian cat behavior and can provide new insights into cat domestication.

W139: Cat & Dog Workshop

Shared Genetics of Obsessive Compulsive Disorder in Dogs and Humans

Elinor K. Karlsson, University of Massachusetts Medical School, Cambridge, MA

Obsessive-compulsive disorder (OCD) is a heritable, severe mental illness manifested as time-consuming repetition of behaviors. In humans, OCD has a highly polygenic architecture that makes it difficult to identify genetic risk factors with currently attainable sample sizes. Canine OCD is a naturally occurring model for human OCD that is genetically more complex than induced animal models. Through single breed genome-wide association and subsequent targeted sequencing in multiple dog breeds, we identify genes and candidate causal variants associated with canine OCD. We then employ a canine and murine model-driven approach to reduce the genetic search space for a human association study, and identify significantly associated genes despite a modest cohort size. Our results implicate cell adhesion molecules involved in synaptic connectivity and maintenance in both dog and human OCD. Applying this approach to a wider range of complex behavioral disorders shared between dogs and humans could lead to new biological understanding and potential therapeutic targets.

W140: Cat & Dog Workshop

The Feline SNP Array: Features and Utility

Hasan Alhaddad, Kuwait University, Safat, Kuwait

The development of high throughput SNP genotyping technologies has radically changed the genetic dissection of simple and complex traits in human as well as in cats. SNP arrays permit genome-wide population based analyses at very low costs. Eight diverse cat breeds and populations were re-sequenced for SNP discovery in the cat genome. The discovered SNPs were used to develop an Illumina high-density array with ~63,000 SNPs. A comprehensive database composed of over 3000 genotyped cats representing over 40 distinct breeds and populations was used to investigate the array's features and utility. The first aim was to investigate the efficiency of the array in various population-based analyses. As a measure of the efficiency of the array, the following were addressed: (1) SNP remapping to their proper location in the latest feline genome assembly (8.0), (2) SNP genotyping rates in over two thousand cats, (3) SNP Mendelian errors rates using 86 trios, and lastly (4) general population specific summary statistics in 48 distinct cat populations. The second aim was to examine the array's utility and power to detect feline population structure and associations with various traits of interest. The genotypes of 2161 individuals cats (48 populations) were used to determine (1) genetic structure among cat populations, (2) extent of linkage disequilibrium in various populations, and (3) power of the array to reveal regions of association with (dominant and recessive) traits/diseases in small cat breeds and in random bred populations.

W141: Cat & Dog Workshop

Compataive Cytogenomics of Canine and Human Cancers

Matthew Breen, North Carolina State University, Raleigh, NC

We have performed genome wide DNA copy number profiling of over 1,200 tumors representing a variety of common cancers. Each biopsy was obtained from patients with detailed histopathology and clinical follow up. We have identified numerous cytogenomic signatures associated with canine cancer subtypes and are using these as a foundation to offer a more sophisticated means of tumor diagnosis and early detection. We have also begun to define genomic lesions that correlate with patient outcome, leading to the development of new prognostic assays. Through genomic recoding we have compared our canine data to similar platform data from the corresponding cancers in human patients. These data provide strong evidence for a shared pathogenetic origin of several cancers affecting both species. By considering both genomes in such a comparative context, we see that the genomic complexity of some key cancers may be less than what human studies alone have suggested. A brief overview of some of these studies will be presented.

W142: Cat & Dog Workshop

Histological Characterization and Genomic Localization of the Lykoi Breed Hair Variant

Barbara Gandolfi, University of Missouri-Columbia, Columbia, MO

W143: Cattle/Sheep/Goat 1

Ruminant Comparative Genomics as a Tool for Tracing the Formation and Evolution of the Production Traits

Denis M. Larkin[1], Marta Farré Belmonte[1] and Harris A. Lewin[2], (1)Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, United Kingdom, (2)UC Davis, Davis, CA

The unique adaptations, ecological importance, and economical significance have made ruminants a prime target for genome sequencing. In this study we focused on the analysis of chromosomal rearrangements that occurred in the ancestral ruminant genome ~60-35 Mya in order to shed light on the evolution of economically important organs. The genomes of artiodactyls, including eight ruminants, and eight outgroup species were compared. We detected evolutionary breakpoint regions (EBRs) in the ancestral artiodactyl genome and ruminant-specific EBRs. To investigate the potential adaptive value of ancestral EBRs, we studied genes and gene networks affected by genome rearrangements in artiodactyls. Ruminant EBRs are enriched for genes related to immunological function, e.g., they reorganize the MHC Class II cluster, which is involved in adaptive immunity and also reorganize the IgA pathway, involved in the rumen's ability to maintain a wide range of symbiotic bacteria.

W144: Cattle/Sheep/Goat 1

Progress Toward a Low Budget Reference Grade Genome Assembly

Benjamin D. Rosen[1], Derek M. Bickhart[2], Sergey Koren[3], Alex R. Hastie[4], Timothy P.L. Smith[5], Steven G. Schroeder[2], Shawn T. Sullivan[6], Ivan Liachko[7], Joshua N. Burton[7], Brian L. Sayre[8], Heather J Huson[9], George E. Liu[2], Erin E. Connor[2], Tad S. Sonstegard[10], Adam Phillippy[3] and Curtis P. VanTassell[2], (1)ARS, USDA, Beltsville, MD, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)National Human Genome Research Institute, Bethesda, MD, (4)BioNano Genomics, San Diego, CA, (5)USDA, ARS, USMARC, Clay Center, NE, (6)Phase Genomics, Seattle, WA, (7)University of Washington - Department of Genome Sciences, Seattle, WA, (8)Virginia State University, Petersburg, VA, (9)Cornell University, Ithaca, NY, (10)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

Reference quality *de novo* genome assemblies were once solely the domain of large, well-funded genome projects. While next-generation short read technology removed some of the cost barriers, accurate chromosome-scale assembly remains a real challenge. Here we present efforts to *de novo* assemble the goat (*Capra hircus*) genome. Through the combination of single-molecule technologies from Pacific Biosciences (sequencing) and BioNano Genomics (optical mapping) coupled with high-throughput chromosome conformation capture sequencing (Hi-C), an inbred San Clemente goat genome has been sequenced and assembled to a high degree of completeness at a relatively modest cost. Starting with 38 million PacBio reads, we integrated the MinHash Alignment Process (MHAP) with the Celera Assembler (CA) to produce an assembly composed of 3110 contigs with a contig N50 size of 4.7 Mb. This assembly was scaffolded with BioNano genome maps derived from a single IrysChip into 333 scaffolds with an N50 of 23.1 Mb including the complete scaffolding of chromosome 20. Finally, cis-chromosome associations were determined by Hi-C, yielding complete reconstruction of all autosomes into single scaffolds with a final N50 of 91.7 Mb. We hope to demonstrate that our methods are not only cost effective, but improve our ability to annotate challenging genomic regions such as highly repetitive immune gene clusters.

W145: Cattle/Sheep/Goat 1

Dimensionality of Genomic Information in Cattle and Other Species

Ignacy Misztal, University of Georgia, Athens, GA

The dimensionality of the additive information in SNP information can be defined as 1) the number of causative SNP if all causative SNP are identified, 2) the number of effective SNP if only generic SNP are available, 3) the number of independent chromosome segments. Measures 2) and 3) are due to large LD blocks and may be identical, with 3) being a linear function of the effective population size. In practice, the number of effective SNP can be computed from eigenvalue decomposition of SNP BLUP matrix or the genomic relationship matrix when both the number of genotyped individuals and SNP markers are large enough. In particular, with SNP50k, that dimensionality approximately corresponds to the number of the largest eigenvalues explaining about 98% of the matrix variation, with the reminder 2% attributed to noise; the fraction due to noise decreases with more SNP and increases with less SNP quality control. Analyses with large number of genotypes indicate approximately 10k effective SNP in Holstein and Angus, about 4k in commercial pigs, and about 3k in commercial broilers. Small number of effective SNP limit the resolution of GWAS but allow for GBLUP based genomic predictions with arbitrary number of animals at a small cost.

W146: Cattle/Sheep/Goat 1

Increasing Feed Efficiency and Reducing Methane Emissions Using Genomics: An International Approach

Christine F. Baes[1], Angela Cánovas[1], Mike Coffey[2], Erin E. Connor[3], Birgit Gredler[4], Getu Hailu[5], Vern Osborne[5], Jennie Pryce[6], Mehdi Sargolzaei[7], Flavio Schenkel[1], Zhiquan Wang[8], Paul Stothard[9] and Filippo Miglior[1,10], (1)Centre for the Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada, (2)Scotland's Rural College, Easter Bush, United Kingdom, (3)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (4)Qualitas AG, Zug, Switzerland, (5)University of Guelph, Guelph, ON, Canada, (6)Department of Environment and Primary Industries / La Trobe University, Victoria, Australia, (7)Semex Alliance, Saint-Hyacinthe, QC, Canada, (8)University of Alberta, Edmonton, AB, Canada, (9)Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (10)Canadian Dairy Network & University of Guelph, Guelph, ON, Canada

Genomic technology (including SNP arrays and next-generation sequencing) is a powerful driver for the genetic improvement of livestock. Strategic phenotyping of several thousand genotyped animals may now suffice for effective selection within an entire population. Rapid development of new technologies and precision farming allow more precise / automatic measurement of existing or new traits.

In collaboration with partners from Australia, Switzerland, the United Kingdom, and the United States, we present a genomics-based approach to improve feed efficiency and reduce methane emissions in dairy cattle. The foundation of this research is the collection of individual daily feed intake and methane emission data for cows and heifers in Canada, as well as feed efficiency and methane emission data from partner countries. The ultimate outcome of this major research initiative will be routine genetic evaluation services for feed efficiency and methane emission to allow for genetic selection and improvement of these novel traits.

The results of this $10 million project are expected to improve feed efficiency of dairy cattle and reduce the environmental footprint of the dairy industry, in part due to lower methane emissions and reduced land required for feed production.

W147: Cattle/Sheep/Goat 1

Association of Selenocysteine Transfer RNA Fragments with Serum Antibody Response to *Mycoplasma spp.* in Beef Cattle

Eduardo Casas[1], Guohong Cai[1], Larry A. Kuehn[2], Karen B. Register[1], John D. Neill[1] and Tara G. McDaneld[2], (1)USDA, ARS, National Animal Disease Center, Ames, IA, (2)USDA-ARS, U.S. Meat Animal Research Center, Clay Center, NE

The objective was to identify transfer RNA fragments associated with a serum antibody response to *Mycoplasma spp.* in beef cattle. Serum from sixteen beef calves was collected at three points: in summer after calves were born, in fall at weaning, and in the following spring. All sera collected in summer were negative IgG reactive with *Mycoplasma spp.* By the fall, eight animals were seropositive (positive group), while eight remained negative (negative group). By spring, all animals in both groups were positive. Small non-coding RNAs were extracted and sequenced on the Illumina HiSeq next-generation sequencer. Based on prototypical features of the transfer RNAs, a total of 261,502,003 sequences were identified as 5' transfer RNA fragments (tRF5), and were further characterized. Sera collected in the spring from the positive group had 4.8 times more selenocysteine tRF5 sequences when compared to sera collected in the same time point from the negative group (P= 0.0135). Selenocysteine is a rarely used amino acid that is incorporated into proteins by the opal stop codon (UGA), and its function is not well understood. Production of selenocysteine tRF5 may be associated to a host defense mechanism triggered by bacterial infection, or it may provide some advantage to a pathogen during infection of a host. Further studies are needed to establish if selenocysteine tRF5 could be used as a diagnostic indicator of chronic exposure, or whether intervention strategies could be developed to be used as an alternative to antibiotics for controlling disease due to *Mycoplasma spp.*

W148: Cattle/Sheep/Goat 1

Genotyping with Arrays and Targeted GBS – Doing More with Less Expense

Heather Koshinsky, Affymetrix, Santa Clara, CA, Curtis P. Van Tassell, USDA-ARS-AGIL, Beltsville, MD and Graham F. Alder, Beef + Lamb New Zealand Genetics, Dunedin, New Zealand

Rapid access to affordable genotype information is a cornerstone of increased revenue through managing genetics. This applies to all species: cattle, goats, sheep, fish, poultry, soy, wheat, oil palm, blueberries, etc. The number of markers of genotype information needed depends on the application. While parentage assignment or seed lot verification require tens to hundreds of markers, marker assisted management or marker assisted breeding requires a few thousand markers and full characterization requires tens to hundreds of thousands of markers. Example technologies to obtain genotype information include low to mid marker plex Eureka™ Genotyping and mid to high plex Axiom® arrays. When sires can be accurately assigned based on a few hundred markers, multi-sire mating may be practical. As the genetic diversity of a species becomes well characterized, new marker discovery saturates, and there is dense genotype information on high influence individuals (plants or animals). In this scenario the genotype of a few thousand well-chosen markers on a single sample can be used to create the imputed dense genotype of that sample. Breeding and /or management decisions may be informed by imputed dense genotypes obtained at the cost of the genotypes of a few thousand markers. It is expected that increasing the number of samples being tested decreases the per sample cost of obtaining the genotype, especially when tens to hundreds of thousands of samples are committed by an organization or association. The access to rapid affordable genotype information will allow agriculture to do more with less expense.

W149: Cattle/Sheep/Goat 1

Annotating the Bovine Genome with Single-Molecule Transcript Sequencing

Christine G. Elsik[1,2], Darren E. Hagen[1], Tara G. McDaneld[3] and Timothy P.L. Smith[3], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)MU Informatics Institute, University of Missouri, Columbia, MO, (3)USDA-ARS, U.S. Meat Animal Research Center, Clay Center, NE

Bovine genome annotation with RNAseq has been challenging due to the complexity of alternative splicing, potential for read mapping errors, high repeat content of the genome and expressed sequences, and the need to assemble full-length transcripts from short reads. PacBio Iso-Seq technology for generating long-read transcript sequences helps us overcome several of these issues. Advantages of Iso-Seq are that long read lengths allow us to map transcript sequences that have repeats and to avoid predicting only partial genes. We report our approach and results of annotating the *Bos taurus* genome using Iso-Seq data generated from seven tissues (adipose, cerebral cortex, liver, lung, muscle, testis and thalamus). We used the long reads to update both the Ensembl and RefSeq gene sets. Changes to the gene sets included the addition of isoforms, extension of coding regions, UTR extensions, and correction of predicted genes that needed to be merged. In addition to updating existing gene sets, we identified and characterized novel gene loci.

W150: Cattle/Sheep/Goat 2

Linking Genomic and Phenotypic Variation in Cattle Using Intermediate Phenotypes

Ben Hayes[1], Hans D. Daetwyler[2], Amanda J. Chamberlain[3], Kathryn Kemper[4] and Mike Goddard[4], (1)AgriBio, Bundoora, VIC, VT, Australia, (2)Department of Economic Development, Jobs, Transport & Resources, Bundoora, Australia, (3)Dairy Futures Cooperative Research Centre, Bundoora, Australia, (4)University of Melbourne, Melbourne, Australia

Identification of causal mutations which affect complex traits in cattle, especially those affecting key traits in dairy and beef production, could improve accuracy of genomic estimated breeding values, particularly across breeds and with greater persistency of accuracy across time. Identification of these causal mutations would also provide insights into the biology underlying such traits. A significant proportion of the genomic variation in cattle, for *Bos Taurus* breeds at least, has been identified. The 1000 bull genomes project now includes whole genome sequences from 1682 cattle of 55 breeds, from which 67.3 million variants (64.8 million SNP, 2.5 million indel) have been identified. Some progress has also been made in identifying structural variants from these genomes. The challenge is now to determine which variants affect complex traits. This challenge is magnified by the fact that the size of effects of causal mutations are likely to be small, given the large number of mutations typically affecting complex traits. One strategy for identifying such causal mutations is to use intermediate phenotypes, such as

gene expression and protein abundance, where mutation effect is much larger than on the complex trait phenotype, Several examples identifying potential causal mutations affecting protein content of milk from dairy cattle are given. The results highlight the need for better annotation of the bovine genome – many of the most significant mutations are in poorly annotated genomic regions, likely regions regulating gene expression. The functional annotation of animal genomes (FAANG) consortium will greatly improve this situation.

W151: Cattle/Sheep/Goat 2

Interactions Between Diet and Rumen Transcriptomic Pathways and Association with Methane Emissions

Ruidong Xiang[1], Suzanne Rowe[2], Arjan Jonker[2], Cesar Pinares-Patino[2], Jody McNally[3], Jude Bond[4], V. Hutton Oddy[4], Phil Vercoe[5], John McEwan[2] and Brian Dalrymple[6], (1)CSIRO Agriculture, St Lucia, Australia, (2)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand, (3)CSIRO Agriculture, Armidale, Australia, (4)DPI NSW, Armidale, Australia, (5)The University of Western Australia, Crawley, Australia, (6)CSIRO Agriculture, St. Lucia, QLD, Australia

Ruminants release substantial amounts of methane (CH_4) from microbial fermentation of plant material, primarily in their rumen, a multilayered forestomach. How the different layers of the rumen respond to diet and influence CH_4 production has not been studied in detail at the molecular level.

A gene expression correlation network was constructed from rumen wall transcriptomes of 24 sheep fed two different amounts and qualities of feed and measured for a range of phenotypes, including CH_4 production.

The network contained two major negatively correlated gene sub-networks representing the epithelial and muscular layers of the rumen wall. Few genes from the muscle sub-network responded to diet. In contrast, the expression of cell cycle genes (including *BRCA1*) and a number of metabolic processes, was positively correlated with dry matter intake (DMI), ruminal short chain fatty acid concentrations and CH_4 production. The majority of gene expression and phenotypic variation was explained by feed consumption level. Conversely, feed quality explained the majority of the variation in expression of the epithelium-innate immunity genes (including epithelial regulatory factors *GRHL3* and *OVOL1)*. The expression of this gene set had the strongest, but still weak and negative, relationship with CH_4 corrected for DMI (CH_4 yield). The rumen epithelium appears to be more responsive to diet than the muscular layers. Transcriptomics shows promise for the identification of pathways underlying the interaction between the host and the methane production systems in the rumen, which will contribute to the future reduction of methane per unit of livestock production.

W152: Cattle/Sheep/Goat 2

EpiDB: An Omics Data Resource for Cattle

James E. Koltes, Department of Animal Science, University of Arkansas, Fayetteville, AR, Eric Fritz-Waters, Department of Animal Science, Iowa State University, Ames, IA and James M. Reecy, Iowa State University, Ames, IA

Livestock genomics researchers are generating mountains of next generation sequencing data that must be deposited in a public repository for publication. The NCBI short read archive (SRA) is a repository for many types of sequence data including RNA-seq and various regulatory markings that may mediate epigenetic regulation. The livestock Epigenetics database (EpiDB) is a resource that filters public RNA-seq, small RNA-seq, ChIP-seq and methyl-seq data by species, tissue and sequencer type. Only Illumina data that passes quality control (FASTQC) and is annotated for tissue type is retained for analysis. All metatdata is captured and stored in a MySQL database that is linked to a web portal where data can be queried based on species, data type, and tissue. Users can download metadata and access all sequence data through links to NCBI. RNA-seq data is processed to allele specific expression values that can be used to identify differential splicing or other gene regulatory effects. In addition, standardized expression panels were calculated to identify tissue specific transcripts and relative expression levels. As a proof of principle, we analyzed publically available bovine functional genomics datasets to develop reference expression profiles. These data will allow tissue specific transcripts and expression levels to be generated to allow for the comparison of gene expression levels across species as well as other downstream analyses.

W153: Cattle/Sheep/Goat 2

Evaluation of Host Response in Calves Challenged with *Mannheimia haemolytica* using Expression Proteomics

Aswathy N. Rai[1], Leslie A. Shack[1], Joseph S. Reddy[1], Wes Baumgartner[1], William B. Epperson[1], Ty B. Schmidt[2] and **Bindu Nanduri**[1,3], (1)College of Veterinary Medicine, Mississippi State University, Mississippi State, MS, (2)University of Nebraska-Lincoln, Lincoln, NE, (3)Institute for Genomics, Biocomputing & Biotechnology, Mississippi State, MS

Bovine Respiratory disease complex (BRD) is a multifactorial disease-affecting cattle and frequently characterized by the clinical onset of bronchopneumonia associated with *Mannheimia haemolytica*. Pathology of BRD is complex and requires the understanding of protein pathways altered during disease manifestation. However, mass spectrometry based expression proteomics to identify protein biomarkers for disease stratification or diagnosis remains underutilized in BRD. In order to identify a baseline protein expression associated with BRD, we chose to focus on lung tissue, which is the terminal site of infection. Utilizing lung biopsy samples from a challenge trial to identify proteins associated with BRD would rely only on clinical signs and could have confounding affects, as clinical signs do not correspond with lung pathophysiology. To address this our study design included collection of post-mortem samples from calves challenged with *M. haemolytica* from regions of the lung that showed lesions as well as non-lesion regions based on pathophysiology. We describe the first comprehensive and comparative proteomic analysis of lung tissue samples collected seven days post-challenge with *M. haemolytica* from affected and unaffected lobes within the same bovine lung, using 1D LC ESI MS/MS. This comparative expression analysis identified proteins in lung tissue that are differentially expressed following *M. haemolytica* challenge irrespective of the area of sampling, and can be be potential diagnostic markers.

W154: Cattle/Sheep/Goat 2

Cattle Sex-Specific Recombination Maps and Genetic Control from a Large Pedigree Analysis

Li Ma, Department of Animal and Avian Sciences, University of Maryland, College Park, MD

Meiotic recombination is an essential biological process that generates genetic diversity and ensures proper segregation of chromosomes during meiosis. From a large USDA dairy cattle pedigree with over half million genotyped animals, we extracted 186,927 three-generation families,

identified over 8.5 million maternal and paternal recombination events, and constructed sex-specific recombination maps for 59,309 autosomal SNPs. The recombination map spans for 25.5 Morgans in males and 23.2 Morgans in females, for a total studied region of 2,516 Mb (986 kb/cM in males and 1,085 kb/cM in females). The male map is 10% longer than the female map and the sex difference is most pronounced in the subtelomeric regions. We identified 1,792 male and 1,885 female putative recombination hotspots, with 720 hotspots shared between sexes. These hotspots encompass 3% of the genome but account for 25% of the genome-wide recombination events in both sexes. During the past forty years, males showed a decreasing trend in recombination rate that coincided with the artificial selection for milk production. Sex-specific GWAS analyses identified *PRDM9* and *CPLX1* to have significant effects on genome-wide recombination rate in both sexes. Two novel loci, *NEK9* and *REC114*, were associated with recombination rate in both sexes, whereas three loci, *MSH4*, *SMC3* and *CEP55*, affected recombination rate in females only. Among the multiple *PRDM9* paralogues on the bovine genome, our GWAS of recombination hotspot usage together with linkage analysis identified the *PRDM9* paralogue on chromosome 1 to be associated in the U.S. Holstein data. Given the largest sample size ever reported for such studies, our results reveal new insights into the understanding of cattle and mammalian recombination.

W155: Cattle/Sheep/Goat 2

Genome-Wide CNV Analysis Reveals Variants Associated with Growth Traits in *Bos indicus*

George E. Liu, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

Apart from single nucleotide polymorphism (SNP), copy number variation (CNV) is another important type of genetic variation, which may affect growth traits and play key roles for the production of beef cattle. To date, no genome-wide association study (GWAS) for CNV and body traits in beef cattle has been reported, so the present study aimed to investigate this type of association in one of the most important cattle subspecies: *Bos indicus* (Nellore breed). We have used intensity data from over 700,000 SNP probes across the bovine genome to detect common CNVs in a sample of 2,230 Nellore cattle, and performed GWAS between the detected CNVs and nine growth traits. After filtering for frequency and length, a total of 231 CNVs ranging from 894 bp to 4,855,088 bp were kept and tested as predictors for each growth trait using linear regression analysis with principal components correction. There were 49 significant associations identified among 17 CNVs and seven body traits after false discovery rate correction ($P<0.05$). Among the 17 CNVs, three were significant or marginally significant for all the traits. We have compared the locations of associated CNVs with quantitative trait locus and the RefGene database, and found two sets of 9 CNVs overlapping with either known QTLs or genes, respectively. The gene overlapping with CNV100, *KCNJ12*, is a functional candidate for muscle development and plays critical roles in muscling traits. This study presents the first CNV-based GWAS of growth traits using high density SNP microarray data in cattle. We detected 17 CNVs significantly associated with seven growth traits and one of them (CNV100) may be involved in growth traits through *KCNJ12*.

W156: Cattle/Sheep/Goat 2

From Genomic Selection to Functional Genomics in Beef *Bos indicus*: A Case Study

Jose Fernando Garcia, Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ. Estadual Paulista, Sao Paulo, Brazil

Following the footsteps of dairy cattle genomics, genomic selection (GS) approaches were implemented in *Bos indicus*(Nellore) in Brazil during the past five years. However, the use of high density SNP chip information (Bovine HD Illumina) in a cattle population phenotyped for growth rates and development traits, is allowing scientists to move beyond just the application of the GS approaches, since it became possible to investigate (and partially understand) the genetic structure underneath the same traits using functional strategies based on GWAS, Signatures of Selection, Runs of Homozigozity and CNV profiles. Many traits, such as those related to growth and size, are controlled by few genes of pleiotropic effects, which opens new perspectives for applying these information in breeding strategies. Additionally, the use of haplotypes as genotypes in such functional analysis is revealing a new strategy to identify genomic candidate regions for selection. On top of these advance, the inclusion of whole genome sequence data to better characterize those regions and eventually bring light to the specific mechanisms involved in traits expression.

W157: Cattle/Sheep/Goat 2

From the Bovine Y Chromosome to Male Fertility

Wansheng Liu, Department of Animal Science, Penn State University, University Park, PA

The mammalian Y chromosome is unique by its presence only in males, its involvement in sex determination and male fertility, its poor conservation in gene content among mammalian lineages and its accumulation of Y-specific repeat sequences during evolution. In this presentation, I will use the bovine Y as an example to demonstrate how to get from genotype to phenotype, i.e. from the Y chromosome to male fertility. I will present four sets of data: (I) Gene content of male-specific region (MSY) in the bovine Y. A total of 28 protein-coding genes/families are present, 12 of which are single copy genes and remaining 16 are multicopy gene families. These gene families have a total of ~1270 genes, made the MSY gene density the highest in the bovine genome. (II) Copy number variation (CNV) of Y-linked genes in Holstein bulls. We found that CNVs of PRAMEY, HSFY, and ZNF280BY are associated with testicular size and semen quality. (III) Transcriptome of the bovine Y during testis development. RNA-seq data from the bovine testis tissue at the age of 20 days, 8 months, and 2 years old revealed that 95% of the Y genes/ncRNAs are expressed predominantly in testis and may involve in spermatogenesis and male fertility. A weighted gene co-expression network has been built. The integration of the co-expression network and the known interaction profiles resulted in a dynamic network that was used to identify functional modules/motifs in spermatogenesis based on gene ontology (GO). (IV) Role of PRAMEY in spermiogenesis. The PRAMEY protein has been confirmed to play a role in acrosome formation and fertilization.

W158: Cattle/Sheep/Goat 2

Genome Canada Collaboration Opportunities and Funding for Livestock Genomics

David Bailey, Genome Alberta, Calgary, AB, Canada

Genome Canada is a not-for-profit organization that acts as a catalyst for developing and applying genomics and genomic-based technologies to create economic and social benefits for Canadians. Genome Canada connects ideas and people across public and private sectors to find new uses for genomics, invests in large-scale science and technology to fuel innovation, and translates discoveries into applications, new technologies,

societal impacts and solutions across key sectors of national importance, including health, agriculture, forestry, fisheries & aquaculture, energy, mining, and the environment.
Together with its six regional Genome Centres and other partners, we are collectively known as the Canadian Genomics Enterprise. We also support research projects aimed at studying and analyzing the ethical, environmental, economic, legal and social issues related to genomics research (GE^3LS). In addition, five science & technology (S&T) innovation centres with cutting edge technical capabilities are in place. Projects and S&T innovation centres are selected based on their international competitiveness and scientific excellence in the framework of Canada's social and economic fabric. To date, we has received $1.1B in funding commitments from the Government of Canada to which has been added over $1.5B in co-funding from other organizations.
Our business model is built on funding and managing large-scale and multidisciplinary, internationally peer-reviewed research projects and science and technology innovation. We welcome the opportunity to collaborate with academia and industry on areas of common importance and thus to leverage our collective resources. We will provide examples of current livestock projects and are open to discussions on future topics.

W159: Cattle/Sheep/Goat 2

Identification of Two Causal Mutations Associated with Milk Fat Content and Milk Fatty Acid Composition in Goats

Pauline M. Martin[1], Cyrielle Maroteau[1,2], Isabelle Palhière[1], Julien Sarry[1], Hüseyin Besir[3], Rachel Rupp[1] and Gwenola Tosser-Klopp[1], (1)INRA-GenPhySE, Castanet Tolosan Cedex, France, (2)Medical research Institute, University of Dundee, Dundee, United Kingdom, (3)EMBL, Heidelberg, Germany

In the framework of the "Phénofinlait" (www.phenofinlait.fr) and the "3SR" (www.3srbreeding.eu) projects, a large daughter design has been carried out in commercial farms for mapping traits in French dairy goats including milk production and fatty acid content estimated by mid-infrared spectra.
2,254 goats and their 20 artificial insemination sires were genotyped with the GoatSNP50 chip. After standard quality controls 49,647 SNPs were validated. QTL detection was performed using the QTLmap software.
Among the numerous QTL detected, a major QTL was found on CHI 14 associated with both the fat content and the fatty acid composition of milk in the region of the DGAT1 gene, which codes for a key enzyme involved in triglyceride synthesis in milk. DGAT1 sequencing in the 20 sires revealed two missense mutations, which were subsequently genotyped on a larger population of AI bucks. The first mutation was found only in the Saanen breed with a frequency of about 3%; the second mutation was present in both breeds with a frequency of 6% and 14% in Alpine and Saanen breeds, respectively. These two mutations seem to be associated with reduction of fat content (First mutation: -0.89g/kg in Saanen; second mutation: -1.19g/kg and -1.31g/kg in Saanen and Alpine respectively, for heterozygous animals), changes in fatty acid composition and increase in protein content. In order to validate the causality of the mutations, the different versions of the protein have been produced by a baculovirus system and first results obtained with CGFID show a reduced activity for mutant recombinant proteins.

W160: Cattle/Sheep/Goat 2

Is Genotyping By Sequencing a Viable Alternative to Existing Methods for Genomic Selection and GWAS?

Shannon Clarke[1], Ken G Dodds[1], Rudiger Brauning[1], Tracey Van Stijn[1], Rayna Anderson[1], Theódór Kristjánsson[2], Suzanne Rowe[1] and John McEwan[1], (1)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)Stofnfiskur hf., 221 Hafnarfjordu, Iceland

Genotyping by sequencing (GBS) is emerging as an alternative technology to array based genotyping for genomic selection and genetic mapping. A method, based on GBS allele read depths, was developed to give unbiased estimates of genomic relationship matrices (GRM). This overcomes issues with missing genotypes and poor genotype calling accuracy due to low coverage when SNP density and samples numbers per lane are maximised. The GRM produced can be interrogated to estimate: breed composition, pedigree, traceability, inbreeding and co-ancestry as well as be included directly in existing mixed models to estimate breeding values. Crucially, the methodology is extremely cost competitive compared to array based technologies, particularly at high densities, which typically involve low numbers of samples and high set-up costs. This opens up major opportunities for application in minor agricultural species and studies of natural populations. Imputation, genome sequence or reference imputation data are not required and the method can be extended to multi-marker/haplotype based methods if an assembled genome is available. Furthermore, with genotyping costs comparable to existing DNA based parentage assays, opportunities are arising to genotype multiple tiers in elite breeding schemes leading to rapid increases in genetic gain.

W161: Cattle/Sheep/Goat 2

Fast Single-Pass Alignment and Variant Calling Using Sequencing Data

Paul M. VanRaden, Animal Genomics and Improvement Laboratory, ARS-USDA, Beltsville, MD and Derek M. Bickhart, Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD

Sequencing research requires efficient computation. Few programs use already known information about DNA variants when aligning sequence data to the reference map. New program findmap.f90 reads the previous variant list before aligning sequence, calling variant alleles, and summing the allele counts for each DNA source in a single pass. Advantages are faster processing, more precise alignment, more useful data summaries, more compact output, and fewer steps. Programs findmap and BWA were compared using simulated paired end reads of length 150 from fragments of length 1,000 at random locations within the UMD3.1 bovine reference assembly. Each base had 1% probability of error and 1% probability of missing. The 39 million variants from run 5 of the 1,000 bull genomes project were included, with every other variant set to reference or alternate. With 1 processor, BWA required 629 minutes per 1X for alignment, whereas findmap required 12 minutes per 1X for alignment and variant calling. Percentage of correctly mapped reads was 90.5% from BWA and 92.9% from findmap. Variant calls were output by findmap only for the 88.2% of pairs where both ends were located within the fragment length and of opposite orientation. Percentages of variants called correctly were 99.8% for SNPs and 99.9% for deletions, while insertions had 99.9% of alternate calls correct but only 98.6% of reference calls. Memory required by BWA was 4.6 Gbytes / processor, whereas findmap required 46 Gbytes that could be shared by multiple processors. Simultaneous alignment and variant calling is an efficient and accurate strategy.

W162: Cattle/Swine

Accelerated Breeding with Non-meiotic Allele Introgression

Scott Fahrenkrug, Recombinetics, Inc, Saint Paul, MN

W163: Cattle/Swine

SNP Introgression by Genome Editing Technology in Pigs

C. Bruce Whitelaw, Roslin Institute, Edinburgh, United Kingdom

Man has been breeding animals for millennia, initially on behavioural characteristics, then visual traits and very recently by selection based on genetic tools. With the unrelenting progress of genomic resources the animal breeder can now utilises sequence variation across the entire genome. Selective breeding exploits the variation that exists within a population of animals, capturing beneficial genotypes to deliver genetic improvement to that population. This process often requires multiple mating regimes and is therefore time consuming. It also relies on the presence or spontaneous appearance of the desired genetic variation in the target breeding gene pool. If the genetic variation does not exist in a given breeding population then it cannot be selected. Neither has it been possible to exploit potentially beneficial variation from other species or other genera because of mating barriers, while expanding the frequency of rare alleles in a population is challenging. The recent development of genome editing technology now enables the use of genetic variation that is not present within a given breeding population. Genome editors are engineered nucleases that direct double-strand breaks to specific target DNA sequences. These tools enable sequence insertions/deletions at the target locus or the introduction of new alleles. Thus, we can now achieve in a single-generation, introgression of a desired haplotype, even from another species, without introducing other unwanted DNA sequences. This opens unprecedented opportunities for agriculture.

W164: Cattle/Swine

The Role of *CD163* in the Sensitivity to Porcine Reproductive and Respiratory Syndrome Virus.

Randall S Prather, University of Missouri, Columbia, Columbia, MO

RS Prather[1], KD Wells[1], KM Whitworth[1], AJ Mileham[2]

[1]University of Missouri, Columbia, MO, [2]Genus plc, DeForest, WI

A recent model of porcine reproductive and respiratory syndrome virus (PRRSV) infection involves the virus entering and infecting porcine alveolar macrophages (PAMs). The first contact between PRRSV and the PAMs had been thought to be via heparan sulfate, and this binding was followed by binding to SIGLEC1. The SIGLEC1-virus complex was then thought to be endocytosed. Within the endosome the virus then was uncoated by CD163. This uncoating released the viral genome to infect the cell. To determine which molecules are involved it is necessary to genetically modify the candidate genes. Initially, we used classical homologous recombination to knockout *SIGLEC1* in pigs. The *SIGLEC1-/-* piglets had no resistance to a PRRSV challenge. To modify the next candidate gene we used a meganuclease (CRISPR/Cas9). In our hands the CRISPR/Cas9 injection into zygotes has proven to be very effective. We have to date edited 5 different genes in piglets by zygote injection. Of the 31 piglets born, 29 were edited. This editing technology is not only easy to use, but highly efficient. We used the CRISPR/Cas9 system to both modify somatic cells prior to somatic cell nuclear transfer and to edit *CD163* in zygotes. The response of *CD163-/-* piglets to a PRRSV challenge will validate or invalidate the original model of the method of PRRSV infectivity.

Funding from Genus plc, NIH (U42OD011140), and Food for the 21st Century.

W165: Cattle/Swine

Gene Editing: Breeding or Genetic Engineering?

Alison Van Eenennaam, University of California, Davis, Davis, CA

The FDA defines “genetically engineered animals” as those animals modified by rDNA techniques, including the entire lineage of animals that contain the modification, and regulates them under the new animal drug provisions of the federal Food, Drug and Cosmetics Act. In this context, the recombinant DNA (rDNA) construct is the new drug, not the animal itself. It is unclear whether the use of new gene editing technologies, such as CRISPRs, in animals will trigger FDA regulatory oversight. Is it the use of rDNA techniques in the *development* of a new genotype, or the *presence* of an rDNA construct in the genome of an animal that triggers regulatory oversight? In 1992 the Office of Science and Technology Policy wrote that regulatory oversight of biotechnology products should not turn on the fact that an organism has been modified by a particular process or technique but rather, “oversight will be exercised only where the risk posed by the introduction is unreasonable, that is, when the value of the reduction in risk obtained by additional oversight is greater than the cost thereby imposed." It is not evident what unreasonable or even unique risks are posed by animals that carry a new genotype (or possibly a genotype/allele that already exists in a species) produced using gene editing techniques as compared to those that occur spontaneously and form the basis of all animal breeding programs. Given the importance of enabling safe innovation, there is an urgent need to determine the appropriate regulatory framework for gene editing.

W166: Cattle/Swine

NRSP-8 Bioinformatics Coordinator Report

James M. Reecy, Iowa State University, Ames, IA

W167: Citrus Genome

Identification of Citrus Species-Specific Genes with RNA-Seq

Javier Terol, Francisco R Tadeo and Manuel Talon, Centro de Genomica, IVIA, Moncada, Valencia, Spain

RNA-seq was used to characterize the transcriptome of vegetative and reproductive tissues from 12 citrus species: *C. medica*, *C. aurantifolia*, *C. limon*, *C. bergamia*, *C. clementina*, *C. deliciosa*, *C. reshni*, *C. maxima*, *C. paradisi*, *C. aurantium*, *C. sinensis*, and *Poncirus trifoliata*. Four different organs were analyzed: root, bark, leaf, and flower.

28 samples were used for RNA-Seq analysis, and total of 3421 million Illumina reads were produced and mapped against the reference *C. clementina* genome sequence. In order to identify new species-specific transcripts the non-mapping reads were de novo assembled using the Genome Workbench software. One assembly was carried out for each species.
Overall 202,625,127 Ilumina reads were assembled, affording 1.41 Gb of sequence. Assembly of reads per species yielded a total of 214,458 contigs, with an average of 19,496 contigs per species. 98,916 contigs that yielded no hits when used as queries against the *C. clementina* reference sequence were further analyzed. A BLASTX search was carried out in order to identify contaminant sequences, resulting in 3.954 contigs that display significant similarity with respect plant species, while the remaining ones were clearly similar to proteins from microorganisms or animals.
The plant contigs consensus sequences were reassembled with the Staden software and a total of 687 supercontigs were obtained, while 1559 consensus remained as singlets. Functional annotation with Blast2GO and further characterization of these transcripts will be reported.

W168: Citrus Genome

Transcriptome-Based Elucidation of Molecular Mechanism Underlying Cold Tolerance of Trifoliate Orange (*Poncirus trifoliata* (L.) Raf.) and Exploration of Cold-Responsive Genes

Ji-Hong Liu, Huazhong Agricultural University, Wuhan, China

Trifoliate orange (*Poncirustrifoliata* (L.) Raf.)is extremely cold hardy after a full acclimation; however the underlying molecular mechanisms underlying this plant remain poorly understood. In this study, we first generate a transcriptome of trifoliate orange and then perform global transcriptome profilesin response to cold (4^{o}C) by high-throughput sequencing. Sequences derived from cold-treated and control plants were mapped to the assembled transcriptome, resulting in the identification of 5,549 differentially expressed genes (DEGs). These comprised 600 (462 up-regulated, 138 down-regulated), 2,346 (1,631 up-regulated, 715 down-regulated), and 5,177 (2,702 up-regulated, 2,475 down-regulated) genes from the cold-treated samples at 6, 24 and 72 h, respectively, when compared with the control. Plant hormone signal transduction, plant-pathogen interaction, and secondary metabolism were the most significantly enriched GO categories amongstin the DEGs. A total of 60 transcription factors were shown to be cold responsive. In addition, a number of genes involved in the catabolism and signaling of hormones, such as abscisic acid and ethylene, were affected by the cold stress. Meanwhile, levels of polyamine were progressively increasedunder cold, consistent with up-regulation of an arginine decarboxylase gene. We further characterize the function of one of these cold-responsive transcription factors using transgenic approach.

W169: Citrus Genome

Genomic Selection in Citrus Breeding: Accuracy of Genomic Prediction

Mai Minamikawa[1], Keisuke Nonaka[2], Eli Kaminuma[3], Hiromi Kajiya-Kanegae[1], Akio Onogi[1], Shingo Goto[4], Terutaka Yoshioka[4], Atsushi Imai[2], Atsushi Toyoda[3], Asao Fujiyama[3], Takeshi Hayashi[5], Yasukazu Nakamura[3], Tokurou Shimizu[4] and **Hiroyoshi Iwata**[1], (1)The University of Tokyo, Bunkyo, Tokyo, Japan, (2)NARO Institute of Fruit Tree Science, Minami Shimabara, Nagasaki, Japan, (3)National Institute of Genetics, ROIS, Mishima, Shizuoka, Japan, (4)NARO Institute of Fruit Tree Science, Shimizu, Shizuoka, Japan, (5)NARO Agricultural Research Center (NARC), Tsukuba, Ibaraki, Japan

Cross breeding of citrus is strongly hindered by their long lifespan, large plant size and extended juvenile phase for seedlings. Genomic selection is an attractive technology to surmount the problems because it enables selection without field-testing. In this study, we evaluated the accuracy of genomic prediction using 106 citrus varieties. Linkage disequilibrium in the varieties extended around 20 Mb. The accuracy of genomic prediction with four linear and two non-linear models was assessed via cross-validation in 17 traits. Correlation coefficient between observed and predicted genetic values was high (> 0.7) in fruit weight, fruit hardiness, ease of peeling, fruit color and fruit flesh color. Linear and Gaussian kernel regression models showed stable accuracy over the traits. To assess the potential of the population for detecting loci controlling traits, we performed genome-wide association study (GWAS). GWAS detected significant associations in all the traits, suggesting the existence of major QTL controlling the traits. However, genomic prediction with the model treating significant SNPs as fixed effects and the remaining SNPs as random effects was less accurate than genomic prediction with the model treating all SNPs as random effects in most traits, suggesting the robustness of genomic prediction based on the whole-genome markers. We assessed the prediction accuracy in the tails of the distribution of fruit size, and found linear kernel was more accurate than Gaussian kernel in the tails. The results suggest that genomic selection is promising in citrus breeding and that linear kernel regression is recommended as a model for genomic prediction.

W170: Citrus Genome

QTL Analysis for Fruit Traits in Citrus

Marcos Antonio Machado, Centro de Citricultura Sylvio Moreira, IAC, Cordeiropolis, SP, Brazil

QTL analysis for fruit traits in Citrus

Maiara Curtolo[1,2], Mariangela Cristofani-Yaly[1], Marco A. Takita[1], Marinês Bastianel[1], Valdenice Moreira Novelli[1], Rodrigo Gazaffi[3], Andrzej Kilian[4] Antonio Vargas de Oliveira Figueira[2] and Marcos Antonio Machado[1], [(1)]Centro de Citricultura Sylvio Moreira, IAC, Cordeiropolis, SP, Brazil [(2)]Centro de Energia Nuclear na Agricultura/USP, Piracicaba, SP, Brazil [(3)]Universidade Federal de São Carlos, [(4)] Diversity Arrays, Technology Pty Ltd, PO Box 7141, Yarralumla, ACT 2600, Australia

This study aimed to construct an integrated genetic map of tangor Murcott and Pera sweet orange, using the DArT_seq^{TM} molecular markers and localize QTLs for twelve fruit traits. A total of 278 F_1 hybrids were genotyped using the DArT-seq^{TM} markers. To build the integrated map we used the *OneMap* program and considered all DArT loci that showed no segregation deviation. The likelihood ratio was used for formation of linkage groups besides the genomic information obtained from one of the available *Citrus sinensis* genome sequence (http://citrus.hzau.edu.cn/orange/index.php), then proceeded to the preliminary ordering of markers to remove redundancy. The integrated map was composed of 932 markers, 160 that segregated in the 3:1 ratio, 394 that segregated for Murcott tangor and 378 that segregated for Pera sweet orange. The markers were linked in 9 groups with genomic coverage of 2,895.51 cM. The map is saturated and perfectly represent the number of chromosomes of the haploid species. According to the analyzes using the "Multiple QTL Mapping" (MQM) and the results of the permutation test,19 QTLs were identified in total, taking into account the fruits characteristics analyzed: diameter (mm), height (mm), ratio of

H/D, weight (g), rind thickness (mm), segments per fruit, soluble solids (%), Acids (%), Juice content (%), number of seeds, ratio of solids/acids and number of fruits per box. The genome (pseudo chromosomes) of *Citrus sinensis* L. Osbeck was compared to the genetic maps and synteny were clearly identified. Further analysis of the regions with the highest LOD scores was carried and we verified the presence of genes that could be associated with the characteristics.
Financial support: FAPESP (08/57909-2) and CNPq (573848/08-4).

W171: Citrus Genome

Huanglongbing (Citrus Greening Disease) and Pathways toward Genetic Management of the Disease

Fred G. Gmitter, University of Florida, IFAS-CREC, Lake Alfred, FL

Citrus Huanglongbing (HLB or citrus greening disease), associated with a phloem-limited, insect-vectored bacterium *Candidatus* Liberibacter asiaticus, has decimated citrus production where ever it has been found in the world. In Florida alone, citrus production has been reduced to one-third of what it was in just 10 years since the disease was first found. The manifestation of disease is complex, as a consequence of perturbations in many aspects of plant metabolism, anatomy and physiology. There are no citrus accessions or sexually compatible germplasm resources that confer resistance to the disease, but there is a tremendous degree of variation in types and severity of symptoms and tree responses. The most sensitive citrus types decline severely and frequently die, while more tolerant types appear to overcome some of the deleterious effects of disease, and some even remain productive regardless of the infection. Genomic tools have been exploited to understand the nature of the host-pathogen interactions in these diverse citrus accessions. Transcriptomic, proteomic, metabolomic and bioinformatic approaches have shed light on some fundamental differences among diverse citrus accessions; these studies will be reviewed, and possible pathways toward genetic management of HLB will be highlighted.

W172: Climate Change and ICRCGC 1

Combating Climate Change: Call of the Century (C5)

Chittaranjan Kole, Sam Higginbottom Institute of Agriculture, Technology & Sciences, Allahabad, India and Jeffrey D. Ehlers, The Bill and Melinda Gates Foundation, Seattle, WA

International Climate Resilient Crop Genomics Consortium (www.ICRCGC.org) was founded in 2011 to formulate the concepts, strategies, tools and techniques to develop climate-smart crop varieties. Its missions includes utilization of the available genomic resources, particularly in the wild crop relatives, and the advanced tools of genomics, specifically next-generation sequencing. ICRCGC organized seven wroksops during the PAG conference in 2012, 2013, 2014 and 2015; one woekshop in collaboration with Bill and Melinda Gates Foundation during the PAG conference in 2012, and one conference in collaboration with International Society of Crop Science held at Brazil in 2012. During these nine workshops, the deliberations were focused on the basic, 'translational' and participatory research and well as deliberations on future priorities. Meantime, a team of ICRCGC members has published a review entiled 'Application of genomics-assisted breeding for generation of climate resilient crops: progress and prospects' (Front. Plant Sci. 6:563. doi: 10.3389/fpls.2015.00563) as an outcome of the white paper drafted by the ICRCGC members (http://www.icrcgc.org/white_paper.html). In paralell, a number of other platforms with crop or geographic focus have been constituted with similar goals as the ICRCGC. Improved synergy, efficiency and outcomes seems likely if the work of these separate groups can be coordinated. The three workshops of Climate Change and ICRCGC I, II and III during PAG 2016 will aim at forging and fostering collaboration of the crop-wise, trait-wise and region-wise collaboration and interaction among academia, funding agencies and policy makers to architect the C^5 platform.

W173: Climate Change and ICRCGC 1

Climate-Resilient Rice Varietal Development and Targeted Deployment in Asia and Africa

Jauhar Ali, PBGB, International Rice Research Institute, Metro Manila, Philippines

W174: Climate Change and ICRCGC 1

Climate-resilient maize germplasm development and delivery in the tropics: challenges and opportunities

Prasanna Boddupalli, International Maize and Wheat Improvement Center, Nairobi, Kenya

Maize is the key crop for food, feed and nutritional security in sub-Saharan Africa (SSA), Asia and Latin America. Developing and deploying high-yielding, climate-resilient maize (with tolerance to drought, heat, waterlogging, and biotic stresses), is critical for reducing the high risk and vulnerability of the maize-growing smallholders in the tropics to the climate variability. Under the Drought Tolerant Maize for Africa (DTMA) Project, jointly implemented by CIMMYT and IITA, in close collaboration with NARS and private sector institutions in 13 countries in SSA, a total of 184 unique drought-tolerant maize varieties have been released during 2007-2014, with nearly 60% of them being hybrids. These varieties perform as well as or better than the commercial varieties currently available on the market under optimum (no water deficit stress) conditions and out-perform the best commercial checks by at least 25-30% under drought stress and low-input conditions. DTMA has also facilitated production and delivery of about 52,000 tons of seed in 2014 in partnerships with nearly 51 seed companies, benefiting an estimated 5 million African households. The Heat Tolerant Maize for Asia (HTMA) project is another major initiative (that began in 2012), targeted towards development and delivery of heat stress tolerant maize germplasm adapted to Bangladesh, India, Nepal and Pakistan. Under HTMA, 11 heat-stress tolerant CIMMYT-derived maize hybrids have been recently allocated to public and private sector partners in South Asia (and several more in pipeline) for varietal release, scale-up and delivery of the improved seed in the target areas. There are also opportunities to significantly enhance genetic gains in the tropics, through networks of well-managed field-based phenotyping at relevant sites, utilization of modern breeding tools/strategies (including marker-assisted breeding, genomic selection, and doubled haploid (DH) technology), and targeted replacement of several 20+ year old, climate-vulnerable obsolete maize varieties. Increasing rainfed maize productivity and buffering the farming communities from climate-induced variabilities in the climate-vulnerable environments requires more intensive multi-institutional efforts and implementation of innovative policies for effectively integrating improved seed with climate-smart sustainable agronomic practices.

W175: Climate Change and ICRCGC 1

Marker Assisted Genotype by Environment Models for Adapting Wheat to Climate Change

David Gouache[1], Matthieu Bogard[2], Delphine Hourcade[3], Xavier Le Bris[4], Olivier Guillaume[2], Xavier Lacaze[2] and Gustavo de los Campos[5], (1)ARVALIS, Boigneville, France, (2)ARVALIS - Institut du végétal, Baziège, France, (3)Arvalis - Institut du végétal, Baziège, France, (4)ARVALIS - Institut du végétal, La Chapelle Saint-Sauveur, France, (5)Michigan State University, East Lansing, MI

Although the history of agriculture and plant breeding shows that many crops, especially wheat, can be adapted to an incredible range of environments, environmental, especially climatic, variability has always posed problems to geneticists and breeders because it generates genotype by environment interactions. With climate change a reality upon us, the challenge will be for breeding to keep pace with trends in temperature and precipitation. Using two case studies in France, we will illustrate how available tools, namely molecular markers and models of genotype by environment (GxE) interaction, can be leveraged to face this challenge. We will illustrate how a marker based ecophysiological model of wheat phenology may help in designing future wheat varieties with adapted phenologies for the future. We will also show how a genomic prediction model integrating a GxE component may be used to simulate genotype performance in a wider range of climate scenarios than what is permitted by field experiments, and how these results might be analyzed to characterize genotypes for robustness to certain stresses and to climatic variability in general.

W176: Climate Change and ICRCGC 1

Foxtail Millet: A C4 Model Crop with Rich Genetic and Genomic Resources for Enhancing Climate Resilience in Cereals and Bioenergy Grasses

Manoj Prasad, National Institute of Plant Genome Research, NEW DELHI, India

W177: Climate Change and ICRCGC 1

TBA

Alison R Bentley, The John Bingham Laboratory, NIAB, Cambridge, United Kingdom

W178: Climate Change and ICRCGC 1

Breeding and Sustainability of Shrub Willow for Marginal Lands in the Northeast US

Lawrence Smart[1], Fred E. Gouker[1], Craig H. Carlson[1], Eric S. Fabio[1], Chase R. Crowell[1], Christine D. Smart[1], Ran Zhou[2], Felipe R. Montes[3], John E. Carlson[3], Armen R. Kemanian[3] and Stephen DiFazio[2], (1)Cornell University, Geneva, NY, (2)West Virginia University, Morgantown, WV, (3)Pennsylvania State University, University Park, PA

Shrub willow is among the best suited perennial crops for bioenergy feedstock production in the Northeast US due to its high yields, ease of conversion, and overall sustainability. Breeding efforts are aimed at improving yields on typical ag land, strengthening pest and disease resistance, and developing cultivars that can produce viable yields on highly disturbed sites, such as reclaimed mine land. We have developed and characterized association and linkage mapping populations of *Salix purpurea* with the aim of identifying QTL and tightly-linked SNPs for marker-assisted selection. We have adopted genotyping-by-sequencing as a high-density marker system and have mapped putative candidate genes for a number of traits related to yield, pest and disease resistance, and sustainability using the genome sequence and annotation developed by the DOE Joint Genome Institute. We have also initiated a survey of the soil microbiomes associated with willow roots across a range of yield trial sites using high-throughput sequencing of soil DNA samples. We have measured and modeled carbon dioxide and nitrous oxide exchange from land planted with shrub willow, as well as modeled potential yield. Further yield increases, improvements in nutrient use efficiency, and adaptation to increasingly marginal sites facilitated by genomics-assisted breeding will broaden adoption of this highly sustainable alternative to fossil fuels.

W179: Climate Change and ICRCGC 2

Rapid-Cycle Plant Breeding Systems are the Key to Climate Change Adaption in Agriculture

Gary Atlin, Bill & Melinda Gates Foundation, Seattle, WA

Plant breeding is the primary mechanism for adaptation of cropping systems to climate change. Most discussion of its role in adaptation has focused on discovering alleles with large effects on heat and drought tolerance. These are important, but phenology and stress tolerance are usually polygenic in architecture. Adaptation will result from applying modern breeding methods for maximizing genetic gains for quantitative traits. These include rapid breeding cycles, high selection intensity, managed-stress phenotyping, and multi-location testing that adequately samples the target population of environments. Bteeding programs are needed that produce a steady stream of improved varieties, with breeding cycles of no more than five or six years, as well as seed systems that quickly replace varieties. Farmers in the developing world should be using varieties that have been developed in the current climate, or roughly in the last 10 years. Seed systems must aggressively replace obsolete varieties. Commercial seed systems in temperate regions achieve this via competitive seed markets and a high level of trust of farmers in new varieties recommended by seed companies, but the situation is very different in the developing world, where many varieties in widespread use were introduced in the late years of the Green Revolution, and have never been replaced. These varieties were developed in a climate different than today's, placing farmers at risk. The strengthened and accelerated public sector breeding systems needed for climate change adaptation must be supported by freer international exchange of elite germplasm.

W180: Climate Change and ICRCGC 2

Climate Change and Orphan Crops in West Africa

Michael Abberton, International Institute of Tropical Agriculture, Ibadan, Nigeria

Development of climate resilient systems supporting enhanced food security in sub-Saharan requires a focus on crops with currently limited research and development. In a West African context this includes Bambara groundnut and African Yam bean. At the Genetic Resources Centre

(GRC) of IITA we have germplasm collections of these crops and are working in partnership with others to study the diversity in these collections and to underpin pre breeding and breeding.

W181: Climate Change and ICRCGC 2

Climate-Smart Cowpea Breeding for Sub-Saharan Africa

Philip A. Roberts, Department of Nematology, University of California - Riverside, Riverside, CA

W182: Climate Change and ICRCGC 2

Towards the Development of Climate-Smart Crops for Africa

Francesca Stomeo[1], Sita Ghimire[1], Jagger Harvey[1], Nasser Yao[1], Josiah Mutuku[1], Tilly Eldridge[1,2], Josephine Birungi[1] and Appolinaire Djikeng[1], (1)BecA-ILRI Hub, Nairobi, Kenya, (2)John Innes Centre, Norwich, United Kingdom

Climate change represents a serious threat for agricultural productivity worldwide. Smallholder farmers in Sub-Saharan Africa (SSA) are even more vulnerable to climate change given the rain-fed nature of their agricultural systems. The Biosciences eastern and central Africa – International Livestock Research Institute (BecA-ILRI) Hub and its African National Agricultural Research System (NARS) partners have been actively involved in the development of climate-smart agricultural tools for resource-poor farmers and consumers. The BecA initiative underpins high-end research activities with state of the art technologies and capacity building. Some of the current research projects include: the establishment of an integrated genotyping service and support (IGSS) platform which will provide genome sequencing-based genetic profiling of target breeding populations and associated bioinformatics data management, analysis, and decision support services to plant breeders; improvement of tropical forage, Brachiaria spp., for biomass, nutrition and adaptation to drought and low fertility soils; characterizing molecular and biochemical interaction dynamics of plant-virus-vector interactions for the development of new host resistance and management strategies in common bean; and development and redeployment of maize cultivars that are less susceptible to mycotoxin accumulation in specific environments. The current status of these initiatives are presented and discussed with a particular focus on the role played by the BecA-ILRI Hub in developing the capacity of African scientists to solve agricultural problems in the region, and collaborative opportunities to help address key challenges in African agriculture.

W183: Climate Change and ICRCGC 2

Deploying Stress Tolerant Rice Varieties Helps Farmers Cope with Climate Change Adversities

Abdelbagi M. Ismail[1], Uma S. Singh[2], Endang M. Septiningsih[2,3], Damien J. Platten[4], R. K Singh[5] and David J. Mackill[6,7], (1)International Rice Research Institute, Manila, Philippines, (2)International Rice Research Institute, Metro Manila, Philippines, (3)Texas A&M University, College Station, TX, (4)International Rice Research Institute, Los Banos, Philippines, (5)International Rice Research Institute, Makati City, Philippines, (6)University of California-Davis, Davis, CA, (7)UCD, Davis, CA

Floods, drought and salinity are major abiotic stresses constraining rice production in over 40 million hectares of rainfed lowlands in Asia, and their effects are worsening with climate changes, causing warmer temperatures, sea level rise and unpredictable rainfall patterns. This is reflected in lower and unstable productivity, leading to severer poverty and food insecurity. Efforts are needed to develop “climate ready” crops to cope with these adversities and to meet the increasing global demand for food. Salinity limits rice productivity in both irrigated and rainfed areas, causing low yields often below 2 t ha^{-1}. Numerous tolerant varieties were commercializied in recent years, with substantial yield advantges in saline areas. QTLs associated with tolerance were identified and are being combined in new varieties for higher tolerance. Both partial (stagnant, SF, 25-50 cm) and compete (submergence) floods can severely damage rice. Traditional varieties are still prevailing in affected areas, with low yields and poor quality. SF tolerant genotypes with higher yields were recently developed and mapping populations are being analyzed to identify QTLs associated with tolerance. Submergence occurs frequently and causes serious yield losses. The discovery and cloning of the *SUB1* QTL helped in breeding varieties that tolerate submergence for 2 weeks, with no undesirable consequences. *SUB1* effectiveness was validated in farmers’ fields with significant yield advantages. Numerous Sub1 varieties were commercialized and are reaching millions of farmers in Asia. *SUB1* is being combined with tolerances of drought and salinity to provide resilient varieties for coping with climate changes.

W184: Climate Change and ICRCGC 2

Comprehensive Understanding of Crop Phytobiomes is Critical for Meeting the Challenges of Climate Change

Kellye Eversole, Eversole Associates, Bethesda, MD, Jan E. Leach, Colorado State University, Fort Collins, CO and Gwyn A. Beattie, Iowa State University, Ames, IA

To meet the demands of a global human population expected to exceed 9.6 billion by 2055, crop productivity in sustainable agricultural systems must improve considerably in the face of a steadily changing climate and increased biotic and abiotic stressors. Achieving this will require a holistic, systems level approach that will combine a wide range of disciplines including agronomy, physiology, genomics, genetics, breeding, physics, and modeling. While significant progress has been achieved in the past years for many crops in developing and characterizing genetic and genomic resources, we still have a very limited understanding of genotype by environment x management (GxExM) interactions that determine productivity, quality, and the ability to withstand biotic and abiotic stressors. It is becoming increasingly clear that microbes are a key component of the environment and play a critical role in plant health and fitness. Thus, we need a better understanding of the interactions between the plant, microorganisms, soil, and climate, in other words, the phytobiome. There is a large diversity of phytobiomes as it will depend on the region, the crop, the soil, etc. Through the Phytobiomes Initiative (www.phytobiomes.org), a draft roadmap for phytobiomes research and translation has been developed and the establishment of an international, public-private phytobiomes consortium is under way. The draft research roadmap and the structure of the new International Phytobiomes Consortium will be presented.

W185: Climate Change and ICRCGC 3

Toward Production Systems Incorporating Ratooning/Perennial Grain Crops

Andrew H. Paterson[1], T. Stan Cox[2], Wenqian Kong[3] and Pheonah Nabukalu[2], (1)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, (2)The Land Institute, Salina, KS, (3)Plant Genome Mapping Laboratory, Athens, GA
Farmers, both smallholders in developing countries and industrial-scale producers in developed countries, need more options for sustainable intensification of crop production while also increasing support of multiple ecological functions (e.g., topsoil retention) to mitigate or preferably reverse long-term losses of ecological capital associated with row crop agriculture. These needs must be met under more challenging circumstances than those of the first 'Green Revolution', including rising input costs, diminishing water resources, increased climate variability, and degraded soil conditions. The development of perennial grain crops, such as wheat, sorghum, and grain legumes, is a potentially transformative innovation that could be a keystone in solutions to these multiple challenges. Smallholders in particular may benefit from this cropping system, to support their hard-won agricultural skills to feed their families. Deeper roots and longer growing periods mean more harvests, better protected soil and more resilience to fluctuations in rainfall – and amortizing seed and soil preparation/sowing costs over multiple cropping cycles may significantly change production economics, perhaps broadening utilization of hybrids. Genomics tools can speed the process from lab to field. The most drought resistant among the five most important cereal crops and a key dual-use (grain and biomass) crop in regions containing some of the world's most degraded soils, sorghum (*S. bicolor* L.) may be an appropriate 'early adopter of multiple-harvest cropping – however preliminary breeding is in progress on perennial variants of several major crops in anticipation of determining their roles and value in region specific cropping systems.

W186: Climate Change and ICRCGC 3
Crops and systems for Emergent Climates in West Africa
Bill Payne, University of Nevada-Reno, Reno, NV

W187: Climate Change and ICRCGC 3
PhotosynQ: Community-driven plant phenotyping for understanding plant responses to climate change
David Kramer, Michigan State Univeristy, East Lansing, MI

W188: Climate Change and ICRCGC 3
Breeding climate resilient sorghum cultivars
David Jordan, University of Queensland, Warwick, Australia
Australian sorghum production environments are characterised by extreme variation in temperature and rainfall creating enormous challenges for crop improvement specialists. Despite these challenges research combining agronomic and breeding interventions has been able to produce a cropping system that is remarkably productive and stable with productivity gains averaging around 4% per annum for the last 20 years. Under the influence of global warming it seems likely that the variability and extremes experienced by sorghum producers in Australia will become more common in other parts of the world, and that approaches used to improve sorghum in Australia will be applicable to sorghum improvement programs worldwide and potentially provide learnings for other crops. In this presentation we will describe our integrated approach to breeding which uses simulation modelling crop physiology and genomics to develop sorghum cultivars adapted to hotter and drier environments.

W189: Coffee Genomics
Building High Quality Reference Genome Assemblies using PACBio long reads for the Allotetraploid Coffea arabica and its Diploid Ancestral Maternal Species Coffea eugenioides
Marcela Yepes[1], Alvaro Gaitan[2], Marco A. Cristancho[3], Luis Fernando Rivera[3], Juan Carlos Correa[3], Carlos Ernesto Maldonado[2], Carmenza E. Gongora[2], Andres Mauricio Villegas[2], Huver Posada[4], Aleksey Zimin[5], James A. Yorke[5], Keithanne Mockaitis[6] and Herb Aldwinckle[1], (1)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (2)Centro Nacional de Investigaciones de Cafe, CENICAFE, Chinchiná, Colombia, (3)Colombian Center for Bioinformatics and Computational Biology (Bios), Manizales, Caldas, Colombia, (4)Federacion Nacional de Cafeteros de Colombia (FNC)/ Centro Nacional de Investigaciones de Café (CENICAFE), Chinchina, Caldas, Colombia, (5)University of Maryland, College Park, MD, (6)Indiana University, Bloomington, IN
Allopolyploids originate from hybridization between divergent genomes associated with chromosome set doubling. As a consequence, the genomes may undergo a wide range of structural, epigenetic, and functional changes. The world's most widely cultivated coffee species, representing 70% of the coffee market, is the allotetraploid, *Coffea arabica* (2n=4x=44; genome size 1.3 Gb). *C. arabica* evolved through the interspecific hybridization of the ancestors of two diploid *Coffea* species: *Coffea eugenioides* (2n=22, maternal donor, genome size 0.66 Gb) and *C. canephora* (2n=22, paternal donor, genome size 0.71 Gb). Sequencing and assembly of the *C. canephora* genome was published recently, Denoeud *et al.* 2014. Science 345: 1181-1184; genome assembly can be accessed at: http://coffee-genome.org. We report here progress to produce high quality reference assemblies for *C. eugenioides* and *C. arabica* using Pacific BioSciences (PACBio) long reads to enable coffee genetics and genomics of coffee and speed up adaptation of the crop to climate change. Climate change is probably the most severe threat currently facing the coffee industry on the global scale. In recent years, extreme weather events in Central America, Colombia, and Brazil have led to coffee production losses of more than US $2 bn. Of major concern is the very narrow genetic base of cultivated coffee varieties, and therefore the urgent need to develop advanced genomic tools to speed up characterization of *Coffea* diversity in its Center of Origin, Ethiopia, which accounts for 98% of the genetic pool, to help broaden the genetic base of cultivated *C. arabica* and speed up adaptation of the crop to climate change.
This abstract will have an extended time and will be presented by co-authors Marcela Yepes and Marco Cristancho.

W190: Coffee Genomics
High-Throughput Targeted Genotyping of *Coffea arabica* and *Coffea canephora* Using Next Generation Sequencing

Marcio Resende[1], Eveline Caixeta[2], Emilly Ruas Alkimin[3], Tiago Vieira Sousa[3], Marcos D.V. Resende[4], Srikar Chamala[5] and Leandro G Neves[1], (1)RAPiD Genomics LLC, Gainesville, FL, (2)EMBRAPA, Viçosa, Brazil, (3)Federal University of Viçosa, Vicosa, Brazil, (4)EMBRAPA - sucursal, Vicosa, Brazil, (5)RAPiD Genomics, Gainesville, FL
Coffee is an important tropical crop in the world. Among the different species, *C. canephora* and *C. arabica* are the most widely planted. One of the challenges for the breeding and genomic characterization of Coffee, specially *C. arabica*, is the low genetic diversity and complex polyploid nature of its genome. Here, we present the development of a multi-species, genome-wide, high-throughput genotyping platform for Coffee. The strategy is based on the targeted genome capture of 40,000 regions in the Coffee genome followed by next-generation sequencing. These regions were bioinformatically identified to avoid repetitive elements and screen a large number of annotated genes. To capture these regions, we designed probes using a combination of genomic resources, including the *C. canephora* reference genome and assembled unigenes specific to each of the two species. We evaluated the method on 72 samples from *C. canephora* and 72 from *C. arabica*. This population resulted in the discovery of 162,026 SNPs in 27,651 polymorphic probes, with a median of 5 SNPs per probe. From this total, 33,239 SNPs were specific to *C. arabica* and 87,271 SNPs were specific to *C. canephora.* The assay resulted in 3% median missing data, out of which 967 and 40 SNPs were missing in all the individuals of *C. arabica* and *C. canephora*, respectively, indicating the discovery of inter-specific presence and absence (PAV) variants. This assay represents a new tool for the Coffee community that can help future genome assemblies, accelerate breeding, unravel the genetic basis of traits of interest and manage genetic diversity in the species.

W191: Coffee Genomics

Mixed Model to Multiple Harvest Location Trial Applied to Genomic Prediction in *Coffea canephora*

Luis Felipe V. Ferrão[1], Romario G. Ferrão[2], Maria A. G. Ferrão[3], Aymbire Fonseca[3] and Antônio Augusto Franco Garcia[1], (1)University of São Paulo (ESALQ/USP), PIRACICABA, Brazil, (2)Instituto Capixaba de Pesquisa,Assistência Técnica e Extensão Rural, Vitoria, Brazil, (3)Instituto Capixaba de Pesquisa e Assistência Técnica e Extensão Rural / Embrapa Café, Vitoria, Brazil
Genomic Selection (GS) has been studied in several crops and has shown potential to increase the rate of genetic gain and reduce the length of the breeding cycle. Despite the relevance, there is a modest number of reports applied to the genus *Coffea*. Nevertheless, the effective implementation depends on the ability to consider genomic models that represent with adequate reliability the breeding scenario in which the species are inserted. Coffee experimentation, in general, is represented for evaluations in multiple locations and harvests (MET), in order to understand the interaction magnitude and predicting the performance of untested genotypes. Therefore, the main objective of this study was to investigate GS models that accommodate MET modeling. For this, an expansion of the traditional GBLUP was proposed in order to accommodate the interactions in the GS model. Different scenarios that mimic coffee breeding were proposed to assess the predictive ability. In terms of goodness of fit, this approach showed the lowest AIC and BIC values and, consequently, the best goodness of fit. The predictive capacity was measured by cross-validation and, in contrast with the GBLUP, the incorporation of MET modeling showed higher predictive accuracy (on average 10-17% higher) and lower prediction errors. All the genomic analyses were performed using the Genotyping by Sequencing (GBS) approach, which showed a good potential to be used in coffee breeding programs. Thus, in conclusion, the results achieved may be used as a basis for additional studies into the Genus *Coffea* and expanded for other perennial crops that have a similar experimentation design.

W192: Coffee Genomics

Threats of Climate Change on Arabica Coffee (*Coffea arabica* L.) in its Center of Origin Ethiopia

Girma Adugna, Gezahegn B Yadessa and Fikre L Ocho, Jimma University, Jimma, Ethiopia
Arabica coffee (*Coffea arabica* L.) contributes 70% of the world's coffee bean production and consumption. Because *C. arabica* evolved in the moist evergreen afromontane rain forests of Ethiopia, it is a remarkably climate-sensitive species. Over the past decade, documented evidence indicates that climate variables, mainly scarce rainfall, increased drought, and increasing temperatures cause major detrimental effects to Arabica coffee production/yield and quality, and ultimately threaten existence of the crop in its center of origin, Ethiopia, the major reservoir of genetic diversity for the species. The direct impact of climate change includes stressed growth of the coffee tree, limited flowering and berry development leading to poor yield and unacceptable quality. Emergence and/or resurgence of severe outbreaks of diseases (leaf rust, coffee berry disease, wilt, and leaf blight), insects (coffee berry borer, antestia bug, leaf miners, scales and aphids) and nematodes are inevitable. The current global areas of coffee production are projected to shrink by 9.5, 17, and 33% in 2020, 2050 and 2070, respectively. Moreover, the future distribution of indigenous Arabica coffee in Ethiopia is forecasted to decline by about 65% in a number of bioclimatically suitable locations, and in the worst scenarios 100% reduction by 2080. Climate change is inevitably threatening the world coffee industry and unique Arabica coffee genetic resources in Ethiopia, unless adaptation and mitigation strategies are collectively implemented soon. Development of advanced genomic tools to accelerate diversity characterization and their enhanced utilization for genetic improvement to generate drought/stress-tolerant, disease- and insect-resistant coffee varieties are major priorities.

W193: Coffee Genomics

Molecular markers reveal high variability among populations of *Coffea arabica* in its native range of the Afromontane forests of Ethiopia

Kassahun Tesfaye[1], Ludo Muller[2], Kim Govers[3], Endashaw Bekele[1] and Thomas Borsch[4], (1)Addis Ababa University, Addis Ababa, Ethiopia, (2)Institut für Biologie/Botanik Freie Universität Berlin, Berlin-Berlin, Germany, (3)Botanischer Garten und Botanisches Museum Berlin-Dahlem and Institut für Biologie/Botanik, Freie Universität Berlin, Germany, Berlin, Germany, (4)Botanischer Garten und Botanisches Museum Berlin-Dahlem and Institut für Biologie/Botanik, Freie Universität Berlin, , Germany., Berlin, Germany
Coffea arabica L. is the only tetraploid species (2x = 44) of the genus *Coffea*, and is the most widely cultivated and traded. The southwestern forests of Ethiopia are its native habitat, where domestication began. Forest (wild), semi-forest (semi-wild), garden, and plantation coffee are the major conventional production systems in Ethiopia, whereby subsequent selection of coffee from wild populations has led to the formation of

numerous landraces (farmer's varieties) and cultivars. Exploring the genetic diversity of Arabica coffee populations in its natural range is an important parameter for conservation and sustainable use. A total of 10 ISSR and 14 different AFLP primer combinations were used to analyze 9 wild populations and several cultivated genotypes; these markers selectively amplified 84 and 565 fragments, respectively. The NJ and UPGMA clustering analysis of 125 *C. arabica*individuals using 84 polymorphic ISSR markers clearly separated wild genotypes from landrace/cultivars and underscored the existence of wild coffee distinct from the semi-domesticated genotypes. Diversity measure using Shannon's index showed various levels of variability within wild populations in Ethiopia. Those in Yayu (0.47), Bonga (0.46), and Berhane Kontir (0.41) showed the highest diversity. Furthermore, AFLP markers detected moderate to high polymorphism (68% - 92%) with overall values of 73.5% (415 loci) among 130 *C. arabica* accessions. Overall both molecular markers clearly revealed the presence of vast genetic diversity within wild and cultivated coffee landraces and this warrants the need for a multi-site *in situ* conservation approach. Development of advanced genomic tools for diversity characterization should accelerate future conservation efforts.

W194: Comparative Genomics

Genomic Diversity of Native Switchgrass Populations in the United States as Revealed by Exome Capture Sequencing

C. Robin Buell[1], Joseph Evans[1], Brieanne Vaillancourt[1], Emily Crisovan[1], Kerrie W. Barry[2], Chris Daum[3], Jerry Jenkins[4], Govindarajan Kunde-Ramamoorthy[2], Rachael Hume[5], Aruna Nandety[6], Ananta Acharya[7], Chew Yee Ngan[2], Qingzhen Jiang[8], Chia-Lin Wei[2], Jeremy Schmutz[4], E. Charles Brummer[9], Shawn Kaeppler[10] and Michael Casler[11], (1)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)Joint Genome Institute, Walnut Creek, CA, (4)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (5)Michigan State University, East Lansing, MI, (6)Noble Foundation, Ardmore, OK, (7)University of Georgia, Athens, GA, (8)The Samuel Roberts Noble Foundation, Ardmore, OK, (9)University of California, Davis, Davis, CA, (10)Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI, (11)USDA Dairy Forage Research Center, Madison, WI

Switchgrass (*Panicum virgatum*) is a perennial grass native to North and Central America that has been cultivated for forage and soil conservation, and recently as a biofuel production crop. Switchgrass is divided into two ecotypes (lowland and upland) based on habitat, genome ploidy, and phenotype with upland accessions generally accumulating lower biomass than lowland accessions, yet capable of overwintering in low temperatures that cause stand failure in lowland switchgrass. Using an exome capture sequencing approach, we sequenced three population panels of switchgrass (Northern Association, Southern Association, Supplemental Southern Association), representing a total of 1169 individuals across 140 populations. We were able to identify ~1.9 million bi-allelic single nucleotide polymorphic (SNP) loci with coverage in all 1169 individuals. These loci were used to construct a genetic distance dendogram and identify population structure, revealing a total of 10 predicted switchgrass population groups, including a previously un-observed lowland central population group in the Louisiana region, a group in the central mid-west region, and a group in southern Florida. Structural variation was also present within the overall panel with 265,486 total copy number variants (CNVs), which when used in genetic distance estimations, were reflective of SNP-derived genetic relationships. Comparative analyses between upland and lowland populations revealed CNVs restricted to the two ecotypes, indicating a shared loss or duplication event. Additional analyses on genome variation between upland and lowland ecotypes are underway to link genes and alleles with phenotypes relevant to biofuel feedstock production.

W195: Comparative Genomics

How do gene content and gene order change over time? – Answers from Comparison of Closely Related Grasses

Thomas Wicker, University of Zurich, Zurich, Switzerland

W196: Comparative Genomics

Structural Variation and the Genetic Diversity of Cotton

Joshua A. Udall[1], Thiruvarangan Ramaraj[2], Aaron Sharp[1] and Justin T. Page[1], (1)Brigham Young University, Provo, UT, (2)National Center for Genome Resources (NCGR), Santa Fe, NM

Allotetraploid cotton species are a vital source of spinnable fiber for textiles. However, the polyploid nature of the cotton genome complicates analyses while raising many evolutionary questions as to the relationships between homoeologous loci. With new whole-genome re-sequencing data from 34 lines of cotton, representing all tetraploid cotton species, we explore the evolution of the cotton genome with greatly improved resolution. Identifying SNPs and structural variants among these 34 lines and their extant diploid relatives, we clarify phylogenetic relationships among tetraploid species, including newly characterized species, and identify introgression between different species of cultivated cotton. We explore the evolution of homoeologs in the A_T- and D_T-genomes and especially the phenomenon of homoeologous conversion. Homoeologous conversion is rare in cotton, perhaps due to the vast difference in chromosome sizes in the two genomes. Several regions of the genome have been introgressed between *G. hirsutum* and *G. barbadense* resulting in superior cultivars with putative beneficial alleles from both species and new combinations of alleles. The genotypic data resulting from this study provide a valuable resource for cotton researchers and breeders, who can freely access the data online at CottonGen. Continued efforts will expand the CottonGen database with sequence data from more lines of cotton, providing biological insight and access to a wealth of diversity sorely needed in the narrow germplasm of cotton cultivars.

W197: Comparative Genomics

Population-scale functional and structural diversity of the wheat genome revealed by transcriptome and exome sequencing

Matthew J. Hayden[1], Shichen Wang[2], William Rutter[2], Alina Akhunova[2], Yanni Lun[2], Katherine Jordan[2], Wei Wang[2], Kerrie Forrest[3], Tim I Sawbridge[4], Joanna Petkowski[4], Surya Kant[5], Hans D. Daetwyler[6], Fan Shi[7], Pippa Kay[7], Raj Pasam[7], Shiaoman Chao[8], Debbie Wong[7], Josquin Tibbits[9], Ben Hayes[3], Luther Talbert[10], Jorge Dubcovsky[11] and **Eduard Akhunov**[2], (1)Department of Primary Industries Victoria, Bundoora, Australia, (2)Kansas State University, Manhattan, KS, (3)Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia, (4)Department of Environment and Primary Industries,

Bundoora, Australia, (5)School of Applied Systems Biology, La Trobe University, Bundoora, Australia, (6)Biosciences Research, DEDJTR, Bundoora, Australia, (7)School of Applied Systems Biology, La Trobe University, Bandoora, Australia, (8)USDA-ARS, Fargo, ND, (9)Department of Environment and Primary Industries, Bundoora, Victoria, Australia, (10)Montana State University, Bozeman, MT, (11)University of California, Davis, CA
Genomic diversity underlies functional and phenotypic diversity in populations and provides the basis for adaptation to changing environment. Recent analyses of the wheat reference genome suggest that the two rounds of polyploidization and expansion of repetitive DNA space could have facilitated the generation of novel structural and functional variation at a high rate. However, due to the complexity and size of the wheat genome information about the scope of its diversity at the population level remains limited. By characterizing exomes and transcriptomes of the geographically diverse lines we investigated the nature and scale of functional and structural variation in wheat. The analysis of 1,000 wheat exomes is allowing us to create a detailed catalog of SNP and PAV variation and describe its possible effect on gene function. The high-density haplotype map was created and tested as a platform for cross-linking genetic mapping and breeding efforts. Eco-geographic patterns of genomic variation are investigated to identify alleles contributing to adaption and assess their role in breeding. Transcriptome sequencing revealed a significant level of homoeolog-specific expression bias and silencing variation among the accessions of distinct geographic origin. Structural and functional divergence among the wheat genomes was reflected in the distribution of selective sweeps among the homoeologous chromosomes. Analyses of population-scale genomic data suggest that the wheat genome has the potential to generate limitless possibilities for the origin of novel variants that could broaden the adaptive potential of wheat. Our current efforts are directed towards building a genome-wide catalog of these adaptive genomic features

W198: Comparative Genomics

A Comparative Genomics Based Approach to Target Alien Chromosome Specific Genes in Wheat

Vijay K. Tiwari[1], Nidhi Rawat[1], Shichen Wang[1], Parveen Chhuneja[2], Bernd Friebe[1], Jaroslav Dolezel[3], Jesse Poland[1], Eduard Akhunov[1], Sanzhen Liu[1], Robert L. Bowden[4] and Bikram S. Gill[1], (1)Kansas State University, Manhattan, KS, (2)Punjab Agricultural University, Ludhiana, India, (3)Institute of Experimental Botany, Olomouc, Czech Republic, (4)USDA- ARS, Manhattan, KS
The tertiary gene pool of wheat represents an excellent resource for agronomically important genes against biotic and abiotic stress that can be used in wheat improvement. *Aegilops geniculata* is an important species from tertiary gene pool of common wheat and holds tremendous potential for wheat improvement against biotic and abiotic stresses. Chromosome $5M^g$ from *Ae. geniculata* TA10437 has been identified as a source of important resistance genes such as *Lr57, Yr40 and Sr53.* Beside these genes, our results have indicated that this chromosome plays an important role in inducing homoeologous pairing even in presence of the *Ph1* gene of hexaploid bread wheat. To facilitate genomic analysis of this chromosome, we flow-sorted chromosome $5M^g$ from a wheat/*Aegilops geniculata* disomic substitution line [DS$5M^g$ (5D)] and sequenced it using an Illumina HiSeq 2000 system at approximately 50x coverage. Paired-end sequences were assembled and used for structural and functional annotation. Comparative genomics based approaches were used to anchor and order $5M^g$ genes. Generated genomic resources were used to develop SNP markers and these SNPs helped us to delineate the *Lr57* and *Yr40* region on rice and *Ae. tauschii* (5D) genomes. Using GenomeZipper based comparative genomics approach we were able to delineate the *Lr57* and *Yr40* region to a rice BAC on chromosome 12, work is in progress towards identification of candidate genes and their evaluations.

W199: Comparative Genomics

Genome Alignment Spanning Major Poaceae Lineages Reveals Heterogeneous Evolutionary Rates and Alters Inferred Dates for Key Evolutionary Events

Xiyin Wang, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA; School of Life Sciences, North China University of Science and Technology, Tangshan, Hebei, China
Multiple comparisons among genomes can clarify their evolution, speciation and functional innovations. To date, the genome sequences of 8 grasses representing the most economically important Poaceae (grass) clades have been published, and their genomic level comparison is an essential foundation for evolutionary, functional, and translational research. Using a formal and conservative approach, we aligned these genomes. Direct comparison of paralogous gene pairs all duplicated simultaneously reveal striking variation in evolutionary rates among whole genomes, with nucleotide substitution slowest in rice and up to 48% faster in other grasses, adding a new dimension to the value of rice as a grass model. We reconstructed ancestral genome contents for major evolutionary nodes, potentially contributing to understanding grasses' divergence and speciation. Recent fossil evidence suggests revisions of the estimated dates of key evolutionary events, implying that the pan-grass polyploidization occurred ~96 million years ago and could not be related to the Cretaceous–Tertiary mass extinction as previously inferred. Adjusted dating to reflect both updated fossil evidence and lineage-specific evolutionary rates suggested that maize subgenome divergence and maize-sorghum divergence were virtually simultaneous, a coincidence that would be explained if polyploidization directly contributed to speciation. This work lays a solid foundation for Poaceae translational genomics.

W200: Components of Apomixis

Differential Gene Expression in Diploid Sexual, Diploid Apomictic and Triploid Apomictic Species of *Boechera*

Martin P. Schilling, Utah State University, Logan, UT
Among other processes, apomixis in many *Boechera* is characterized by a meiotic 1^{st} division restitution. To better understand the molecular biology of this form of apomeiosis, we focused on RNASeq gene expression data from immature pistils of diploid sexual *B. stricta*, the diploid apomicts *B. lignifera* and *B. retrofracta* x *exilis* and the triploid hybrid apomict *B. gunnisoniana* x *sp*. We specifically addressed the following questions: 1) Can we find a uniform response to drought-stress among the four species in question? 2) Are there genes, associated with apomixis, that can be found across diploid and triploid species of apomictic *Boechera*? and 3) Do triploid *Boechera* apomicts exhibit differences in gene expression, compared to their diploid congenerics? Bayesian estimation based on normalized expression counts was used to identify genes unique to six models, each assuming differential expression between different sets of samples. 476 genes were uniquely up or down-regulated (unique) in apomicts; 408 were unique to a given species; 124 were unique to ploidy (diploid verses triploid); and 36 were unique to diploid or

triploid apomicts compared to diploid sexuals. While evidence of species-specific drought stress responses was clear from principle component analyses, none of the 26072 genes sequenced in our study were expressed in a pattern conforming to a uniform drought stress response across the four species. Differentially-expressed genes included Argonaute 9 (AGO9), upregulated in apomicts by 153-fold; beta glucosidase 18 (AtBG1), upregulated in apomicts by 336-fold; S-locus lectin protein kinase family protein, upregulated in apomicts by 12.7-fold; AGAMOUS-like 48 (AGL48), Domains Rearranged Methylase 1 (DRM1), and SKP1-interacting partner 6 (SKIP6), all up-regulated in apomicts. Down-regulated in apomicts included Prefoldin chaperone subunit family protein (PFDN), Glucose-methanol-choline oxidoreductase family protein (GMCo), histone superfamily protein (H3.1), AGL19, and AGL22. These and others will be discussed.

W201: Components of Apomixis

DNA Methylation Events Specific of Apomictic Lineages in the Boechera Genus

Emidio Albertini[1], Marika Bocchini[1], Manoj Srivastava[2] and John G. Carman[3], (1)University of Perugia, Perugia, Italy, (2)Indian Grassland and Fodder Research Institute, Jhansi, India, (3)Utah State University, Logan, UT

Apomixis (asexual seed formation) could reproductively-stabilize hybrid vigor, but despite many efforts it has not been successfully transferred to crops. Genes controlling apomixis in wild species are still under investigation, and the existence of an epigenetic model has been postulated. Some studies have demonstrated that during sexual reproduction a genome-wide decrease of DNA methylation takes place and is compensated by de novo methylation. In the present study, we investigated DNA methylation in pre-meiotic flower buds of four diploid sexual species, *B. retrofracta* (Utah), *B. stricta* (Utah), and eastern (Cody, Wyoming) and western (Imnaha, Oregon) sexual genotypes of *B. microphylla*, and four diploid apomictic *Boechera* species or species hybrids, *B. lignifera* (Wyoming), *B. retrofracta* x *stricta* (Colorado), *B. retrofracta* x *exilis* (Utah) and *B. microphylla* (Utah). Cytological analyses revealed diplospory in the apomictic genotypes except for *B. microphylla*, which was aposporous. Our aim was to identify epigenetic differences in flower buds between sexual and apomictic species by analyzing DNA methylation. A total of 1203 clear and reproducible bands were amplified. The extent of DNA methylation ranged from 70.55% in apomictic *B. retrofracta* to 83.54% in *B. yellowstonensis*. In particular DNA methylation averaged 77.53% in apomictic species and 81.12% in sexual species. We then checked the entire collection for polymorphisms and found 10 that were absent in all sexual species and present in all apomictic species or *vice versa*. Amplicons were sequenced. The results will be presented and discussed.

W202: Components of Apomixis

Induced Parthenogenesis with an Apomict-Derived Gene

Joann A. Conner, Maricel Podio and **Peggy Ozias-Akins**, University of Georgia, Tifton, GA

Apomixis in P*ennisetum*/*Cenchrus* is the consequence of aposporous embryo sac development and egg cell parthenogenesis, processes controlled by at least two linked genes within the apospory-specific genomic region (ASGR). Parthenogenesis of unreduced eggs occurs in natural apomicts within the genus and backcross lines of pearl millet in which the ASGR-carrier chromosome has been introgressed. Parthenogenesis of reduced eggs is rare in natural and synthetic apomicts, but frequent in sexual pearl millet, rice and maize when transformed with an AP2-domain transcription factor-encoding gene from *Pennisetum squamulatum* (*Cenchrus squamulatus*). Haploid plants can be recovered from parthenogenetically-derived embryos. This gene has potential utility for haploid induction through parthenogenesis in reduced eggs or apomixis through induction of parthenogenesis in unreduced eggs.

W203: Components of Apomixis

A stress-induced polyphenic switch from apomeiosis to meiosis occurs in *Boechera* (Brassicaceae) that is cytologically and molecularly comparable to those of other kingdoms

John G. Carman[1], Mayelyn Mateo de Arias[1], Martin P. Schilling[1], David A. Sherwood[1], Krishna K Dwivedi[2], Manoj K. Srivastava[2] and Shai J. Lawit[3], (1)Utah State University, Logan, UT, (2)Indian Grassland and Fodder Research Institute, Jhansi, India, (3)DuPont Pioneer, Johnston, IA

In many eukaryotic apomicts, stress activates an epigenetic switch from apomixis (unreduced gamete formation and parthenogenesis) to sex (meiotic gamete formation and syngamy). For this to occur, a third polyphenic gender, an apomictic female, is often observed. She is well adapted to favorable conditions, is adept at producing clonal progeny, and, though genetically identical, often differs phenologically and morphologically from sexual females. Stress-activated epigenome reprogramming occurs in her progeny during their development such that sexually-functional males and females form. The genetically-identical siblings often mate to produce fertilized spores or eggs, which tolerate the stresses that induced their formation, e.g., starvation, cold, heat, desiccation, etc. When favorable conditions return, apomictic females form from the fertilized spores or eggs. This cyclical-apomixis pattern occurs in animals, fungi, protists, diatoms, and brown and green algae. In apomictic angiosperms, polyphenic genders are not generally recognized. Instead, apomixis is usually considered to be facultative, with seeds often forming sexually and apomictically on the same plant. Herein, we document for *Boechera* (Brassicaceae) stress-induced switching from mostly apomeiotic to mostly meiotic ovule development, and we show that this switching includes global epigenetics-based changes in gene expression that produce gene ontology patterns in ovaries typical of those observed during reproduction in obligate sexual plants. Our findings unify angiosperms with other eukaryotes in terms of *i*) tendencies for stress-induced polyphenic reversions from apomixis to sex and *ii*) physiological adaptations in apomicts that maximize fecundity thus facilitating geographic parthenogenesis. A primitive origins hypothesis for sex-apomixis polyphenisms, with functionality more-or-less conserved among eukaryotes, is supported.

W204: Components of Apomixis

Male Fertility Genes In Wheat (Triticum aestivum) and Their Use In Seed Production Technology

Marc C. Albertsen, DuPont Pioneer, Johnston, IA

W205: Components of Apomixis

High Incidence of Biallelic Mutation of Floral Genes in CRISPR/Cas9 Transgenic Poplars

Estefania Elorriaga, Amy Leigh Klocko, Cathleen Ma and Steven H. Strauss, Oregon State University, Corvallis, OR

Gene flow from genetically engineered (GE) trees into feral or wild populations are significant obstacles to their use as a result of regulatory, public perception, and ecological concerns. Loss-of-function mutations in a number of floral transcription factor genes can lead to sterility in diverse plant species, however such mutations are rare and generally recessive, thus are very difficult to induce via conventional tree breeding. The recently rediscovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Cas system has proven to be a powerful directed-mutagenesis tool in many species, including trees. The CRISPR-Cas induced mutations appear to be highly predictable and stable, and reversion should be extremely rare or impossible (e.g., when there are deletions of essential parts of coding regions). We are testing the mutation efficiency of six nucleases targeting two essential floral genes in *Populus*. The targets are the poplar orthologs *LEAFY* and *AGAMOUS*—well studied genes essential for both male- and female-fertility. The nucleases have been stably transformed into hundreds of independent events that we are now analyzing for mutation rate and type. Sequencing results from 300 events indicate that the rate of identical mutation in both *LEAFY* alleles is ~32% and in both *AGAMOUS* alleles is ~27%. This high rate of identical biallelic mutations may be due to CRISPR-Cas-mediated allelic conversion during initial mutagenesis of individual callus cells. Results from study of the full population will be presented.

W206: Compositae

Using Phylogenomics to Resolve Mega-Families: An Example from Compositae

Jennifer Mandel, University of Memphis, Memphis, TN, Rebecca Dikow, Smithsonian Institute for Biodiversity Genomics, Washington, DC, DC and Vicki Funk, Smithsonian Institution, Washington, DC

Next-generation sequencing and phylogenomics hold great promise for elucidating complex relationships among large plant families. Here we performed targeted capture of low copy sequences followed by next-generation sequencing on the Illumina platform in the large and diverse angiosperm family Compositae (Asteraceae). The family is monophyletic based on morphology and molecular data, yet many areas of the phylogeny have unresolved polytomies and interpreting phylogenetic patterns has been historically difficult. In order to outline a method and provide a framework and for future phylogenetic studies in the Compositae, we sequenced 23 taxa from across the family in which the relationships were well established as well as a member of the sister family Calyceraceae. We generated nuclear data from 795 loci and assembled chloroplast genomes from off-target capture reads enabling the comparison of nuclear and chloroplast genomes for phylogenetic analyses. We also analyzed multi-copy nuclear genes in our data set using a clustering method during orthology detection, and we applied a network approach to these clusters—analyzing all related locus copies. Using these data we produced hypotheses of phylogenetic relationships employing both a conservative (restricted to only loci with one copy per targeted locus) and a multigene approach (including all copies per targeted locus). The methods and bioinformatics workflow presented here provide a solid foundation for future work aimed at understanding gene family evolution in the Compositae as well as providing a model for phylogenomic analyses in other plant mega-families.

W207: Compositae

Dovetail/ *in vitro* Proximity Ligation Data Facilitates Analysis of an Ancient Whole Genome Triplication Event in *Lactuca sativa*

Sebastian Reyes Chin-Wo, Dean Lavelle, Maria Jose Truco, Alexander Kozik and Richard Michelmore, Genome Center, University of California, Davis, CA

Increases in ploidy level are widespread phenomena in plant genome evolution. Most crops have undergone polyploidization events, some recent and some ancient. A paleoploidization event resulting in a whole genome duplication has been suggested to be basal to all the members of the Compositae (Asteraceae) family based on analysis of duplicated genes. We have investigated this event using a high quality genome assembly and annotation of cultivated lettuce, *Lactuca sativa*. The HiRise program developed by Dovetail Genomics was used to improve the assembly using reads from multiple Chicago libraries, which were generated with an *in vitro* proximity ligation technique. This greatly increased the contiguity and quality of the *L. sativa* genome assembly; the improved assembly was validated genetically. Combined with the curated gene set, this produced a high quality resource for syntenic analysis. Intragenomic syntenic analysis of the *L. sativa* genome and syntenic comparisons with other Asterid species revealed multiple paleoploidization events that occurred basal to the Compositae family. Syntenic segments are often present in triplicate indicative sequential genome duplications rather than the originally proposed single duplication event. Analysis of large segments of the genome allowed precise definition of paralogous groups of genes and analysis of their evolution subsequent to the whole genome duplication events.

W208: Compositae

DNA-Free Genome Editing in Lettuce with Preassembled CRISPR-Cas9 Ribonucleoproteins

Sunghwa Choe, Seoul National University, Seoul, South Korea

As an important fresh vegetable crop, lettuce awaits various trait enhancements. We have sought alternative ways of molecular breeding based on CRISPR/Cas9 technology. Editing plant genomes without introducing foreign DNA into cells may alleviate regulatory concerns related to genetically modified plants. We transfected preassembled complexes of purified Cas9 protein and guide RNA into plant protoplasts of lettuce and achieved targeted mutagenesis in regenerated plants at frequencies of up to 46%. The targeted sites contained germline-transmissible small insertions or deletions that are indistinguishable from naturally occurring genetic variation.

W209: Compositae

One Locus to Rule Them All: Life History QTL in the Silverleaf Sunflower

Brook T. Moyers[1,2], Gregory J. Baute[1] and Loren H. Rieseberg[1], (1)University of British Columbia, Vancouver, BC, Canada, (2)Colorado State University, Fort Collins, CO

The silverleaf sunflower, *Helianthus argophyllus*, is closely related to the domesticated sunflower and has been the donor of agronomically important genotypes for traits including downy mildew resistance and cytoplasmic male sterility. The silverleaf sunflower also exhibits remarkable variation in life history across its native range in coastal South Texas: plants are either short and early flowering with relatively large flowerheads, or tall and late flowering with relatively small flowerheads. Both life history syndromes can be observed in the central, coastal

portion of the species range, while inland populations to the North or South are exclusively late flowering. Experimental crosses show that the early flowering syndrome is dominant over late flowering, and many quantitative traits are strongly genetically correlated with flowering time. We created two replicate F2 mapping populations and evaluated both populations in an agricultural setting in Northern California, measuring days to flower, height at flowering, specific leaf area, height at maturity, branch number, branching degree, basal stem circumference, and basal stem density. Both F2 populations exhibited transgressive values for all traits, but the distribution of traits remained bimodal and most traits were highly correlated with flowering time. We used a *PstI–MspI* genotyping-by-sequencing approach to construct a genetic map and identify quantitative trait loci for each trait. Strikingly, most traits show a dominant single QTL that appears to co-localize with previously identified QTl for flowering time in the sister species *H. annuus* and putatively contains three paralogs of *FT*.

W210: Compositae

Genetic Cloning of Genes Controlling Lettuce Leaf Color

Wenqing Su, Rong Tao, Weiyi Zhang, Zhen Yue, Wenye Liu, Jiongjiong Chen and **Hanhui Kuang**, Huazhong Agricultural University, Wuhan, China

Anthocyanidin not only provides health benefits to consumers but also contributes sensory quality to fruits and vegetables such as lettuce. Some cultivars have leaf colors varying from pink to purple due to the presence of different types and concentration of anthocyanidin. Four leaf color segregating populations of lettuce, including one with red spots, were generated. Surprisingly, lettuce leaf color tends to be quantitative trait, controlled by multiple loci. We used a BSA in combination of RNA-seq approach to dissect the genetic loci controlling lettuce leaf color. CAPS markers were designed for each detected locus, and were used to screen the segregating populations for individuals that are heterozygous at a targeted locus but homozygous at other color-controlling loci. The individuals were self-crossed to produce single-locus segregating population (such as $F_{2:3}$ sub-population), which was used to fine mapping and genetic cloning of genes controlling lettuce leaf color. Consequently, four genes have been successfully cloned, including two MYB, one bHLH and one ANS (anthocyanindin synthase) gene. Interestingly, one of the detected MYB genes is recessive, likely as a competitor of an active MYB gene. Segregating populations with Mendelian ratio for five additional loci have been generated, and fine mapping and cloning of target genes are underway. Furthermore, genes potentially involved in production of anthocyanindin were identified using net-work analysis, and several novel genes were detected. The molecular mechanism for all genes confirmed to be involved in color formation in lettuce will be investigated in future study.

W211: Compositae

Result of the *de novo* Sequencing of the Complex Sunflower Genome Using PacBio Technology (100X)

Jérôme Gouzy[1], Baptiste Mayjonade[1], Christopher J. Grassa[1], Sébastien Carrère[1], Erika Sallet[1], Ludovic Legrand[1], Nicolas Pouilly[1], Marie-Claude Boniface[1], Nicolas Blanchet[1], Brigitte Mangin[1], Navdeep Gill[2], Thuy Nguyen[2], John E. Bowers[3], Cécile Donnadieu[4], Nolan Kane[5], Hélène Bergès[6], John M. Burke[3], Loren H. Rieseberg[2], Stéphane Muños[1] and **Nicolas Langlade**[1], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (2)University of British Columbia, Vancouver, BC, Canada, (3)Department of Plant Biology, University of Georgia, Athens, GA, (4)Plateforme Genomique, Castanet tolosan, France, (5)University of Colorado, Boulder, CO, (6)INRA - CNRGV, Castanet Tolosan, France

The sunflower (Helianthus annuus) genome is large (17 chromosomes spanning 3.6Gb) and complex. It has been shown to be composed of over 81% of transposable elements: mostly recent duplication of Gypsy and Copia Long-Terminal-Repeat elements (Staton et al., 2012). Our strategy using short Roche and Illumina reads from a range of librairy sizes, combined with a high density genetic map and a Finger Printing Contig physical map coud not completely to solve the great complexity of this genome.

In this context, within the SUNRISE project, we started an ambitious approach to generate a high quality genome sequence of the INRA line XRQ, by exploiting the PacBio RSII sequencing progresses. We produced, in majority on the INRA Toulouse sequencer, over 102X of long subreads (MEAN 9.8kb, N50 13.7kb, MAX 80.9kb). After correction using genetic information, the assembly process allowed us to produce 12718 contigs corresponding to 2.93Gb of sequence (MEAN 238kb, N50 524kb, MAX 3 350kb) that were scaffolded using geneitc and physical mapping informations.

In this talk, we will present our approach, the overall assembly process and a first analysis of the genome annotation together with the webportal (www.heliagene.org) functionnalities where we provide this reference genome to the community.

W212: Computational Gene Discovery

Expanding the Repertoire of Small Secretory Peptides in Plants

Yao-Cheng Lin[1], Sarieh Ghorbani[1], Yves Van de Peer[1] and Pierre Hilson[2], (1)VIB - Ghent University, Zwijnaarde, Belgium, (2)Institut Jean-Pierre Bourgin, Versailles, France

Plant genomes encode numerous small secretory peptides (SSPs) whose functions have yet to be explored. Based on structural features that characterize SSP families known to take part in postembryonic development, this comparative genome analysis resulted in the identification of genes coding for oligopeptides potentially involved in cell-to-cell communication. Because genome annotation based on short sequence homology is difficult, the criteria for the *de novo* identification and aggregation of conserved SSP sequences were first benchmarked across five reference plant species. The resulting gene families were then extended to 32 genome sequences, including major crops. The global phylogenetic pattern common to the functionally characterized SSP families suggests that their apparition and expansion coincide with that of the land plants. The SSP families can be searched online for members, sequences and consensus (http://bioinformatics.psb.ugent.be/webtools/PlantSSP/). Looking for putative regulators of root development, *Arabidopsis thaliana* SSP genes were further selected through transcriptome meta-analysis based on their expression at specific stages and in specific cell types in the course of the lateral root formation. As an additional indication that formerly uncharacterized SSPs may control development, this study showed that root growth and branching were altered by the application of synthetic peptides matching conserved SSP motifs, sometimes in very specific ways. The strategy used in the study, combining comparative genomics, transcriptome meta-analysis and peptide functional assays *in planta*, pinpoints factors potentially involved in non-cell-autonomous regulatory mechanisms. A similar approach can be implemented in different species for the study of a wide range of developmental programmes.

W213: Computational Gene Discovery

Automated Functional Annotation of Protein Products of Alternatively Spliced Genes

Michael Sammeth, Institute of Biophysics / Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Alternative splicing (AS) is an important process in gene regulation and substantially contributes to the proteome diversity. Modern genomics employ increasingly high-throughput technologies to obtain the genome and all forms of expressed genes from a studied plant or animal. To overcome difficulties of maintaining AS databases updated with all newly available transcriptome data, we previously introduced AStalavista, a computer program that reports within minutes all AS events on a genome-wide scale.

Here we present AStafunk, an automated pipeline to annotate the functional impact caused by AS. Employing knowledge about protein families, we integrate algorithms for the prediction of functional domains with the efficient detection of AS events. Such studies have hitherto been hampered by difficulties to map the effect of genomic events on the proteome level, which required cumbersome projections between aminoacid and nucleotide coordinates and also implied the inefficient analysis of highly redundant sequences intrinsic to AS transcriptomes.

In a nutshell, the AStafunk software requires as an input (1^{st}) the genomic context, (2^{nd}) gene models and (3^{rd}) protein domain models to provide the exhaustive landscape of AS events and how these are modifying protein domains (i.e., events creating/disrupting or modifying a domain). During the workshop we revise the constrained Dynamic Programming algorithm that limits protein domain search to alternatively spliced regions, with practical applications of the approach for the automated annotation of the functional impact caused by alternative splicing in animal and plant genomes. http://astalavista.sammeth.net

W214: Computational Gene Discovery

Hybrid Method of Using RNA-Seq for Genome Annotation

Alexander Souvorov, NIH/HLM/NCBI, Bethesda, MD

Usually, when RNA-Seq data is used for a genome annotation the reads are aligned to the reference genome, and these alignments are used as building blocks for the reconstructed gene models. The reads also could be assembled into a set of full-length transcripts using a *de novo* transcriptome assembler. In our previous work we demonstrated that a typical draft eukaryotic genome may have anywhere from hundreds to thousands genes for which some exons fall into genomic gaps. The reconstructed transcripts of these genes could be dramatically improved if we fill in the missing exons using appropriate parts of the *de novo* assembled transcripts.

Assembling the full set of RNA-Seq data for a deep transcriptome sequencing project is a computationally expensive process which can yield millions of transcripts. Only a small fraction of these transcripts would be used for the gap-filling of partial genes. For this purpose we propose a hybrid annotation method in which missing parts of the transcripts are assembled on demand only when gene fragments separated by a genomic gap are found. We show that partial transcripts derived from the genome could be used to locate a subgraph of the general de Bruijn graph. This subgraph is usually small enough to be analyzed fully on the fly and produce one or more transcripts connecting the fragments. This method is evaluated on the baboon and rabbit genomes. The baboon gap-filled transcripts are compared with the human data. The rabbit ones are compared with the transcripts produced by Trinity assembler.

W215: Computational Gene Discovery

Orthology-Based Genome Annotation and Interpretation

Robert M. Waterhouse, Massachusetts Institute of Technology & Broad Institute of MIT and Harvard, Cambridge, MA; University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland

The OrthoDB [Kriventseva, *et al*. 2015] catalog of orthologs, www.orthodb.org, 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 [Simão, *et al*. 2015], provide 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 rapidly growing numbers of newly-sequenced plant and animal genomes. Such comparative approaches are well-established as immensely valuable for gene discovery and characterization, helping to build resources to support biological research. Orthology delineation is a cornerstone of comparative genomics, offering evolutionarily-qualified hypotheses on gene function by identifying "equivalent" genes in different species, as well as highlighting shared and unique genes that offer clues to understanding species diversity and provide the means to begin to investigate key biological traits – for both large-scale evolutionary biology research and targeted gene and gene family studies. The success of such interpretative analyses relies on the comprehensiveness and accuracy of the input data, making BUSCO quality assessment an important part of the process of genome sequencing, assembly, and annotation. Orthology-based approaches therefore offer not only a vital means by which to begin to interpret the increasing quantities of genomic data, but also to help prioritize improvements, and to ensure that initial "draft" genomes develop into high-quality resources that benefit the entire research community.

W216: Computational Gene Discovery

Three Methods to Transfer Annotation between Species or Strains

Mario Stanke[1], Stefanie König[1], Simone Lange[1], Ian T. Fiddes[2], Thomas Keane[3] and Lars Romoth[1], (1)University of Greifswald, Greifswald, Germany, (2)University of California, Santa Cruz, CA, (3)Sanger, Cambridge, United Kingdom of Great Britain and Northern Ireland

The accuracy with which gene structures in new target genomes can be predicted can be improved when leveraging a high quality annotation of a related genome. In the talk I will present and compare three conceptually different methods for this task. The first method, TransMap, uses a pairwise genome alignment to map transcript sequences, including alternative splice variants, from a source genome to the target genome, and is followed by small corrections only, optionally supported by target specific RNA-Seq. Another method - comparative AUGUSTUS - uses a multiple genome alignment, a phylogenetic tree, existing annotations for a subset of the genomes, RNA-Seq for another subset of the genomes as well as *ab initio* model components to identify gene structures in all genomes. The gene structures are preferred to agree on aligned exon boundaries. A third method uses spliced alignments of protein sequences, e.g. from GenomeThreader or exonerate, to produce evidence for conserved parts of a gene structure, which is then jointly considered with other evidence by AUGUSTUS. The three approaches aim at different

degrees of similarity between the target genome and the source genome(s) and proteins, respectively. We applied the first two methods to 18 *de novo* assembled mouse strains and tested the last two methods on the *Drosophila* clade.

W217: Computational Gene Discovery

Structural Gene Annotation of Fungal Genomes and Metagenomes

Alexandre Lomsadze, Liexiao Ding and Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA

Accurate gene annotation presents a challenge well known to the fungal genomics community. Structural gene annotation of complete or draft fungal genomes is difficult both in 'short' genomes, e.g. yeast, where the number of introns is too small to parametrize intron models, and in 'long' genomes, where large number of introns per gene presents, in fact, less issue than sequence composition heterogeneity and large populations of transposable elements. Finally, algorithms of structural annotation of fungal metagenomes simply do not exist.

We present here the first, to the best of our knowledge, method for gene identification in fungal metagenomes that takes advantage on availability of hundreds of already sequenced fungal genomes (Ding et al., in preparation). Development of this method was facilitated by experience gained in previous work of the gene prediction tool for prokaryotic metagenomes, MetaGeneMark (Zhu et al, 2010). The new algorithm uses G+C content of input sequence for selecting a model from a set of fungi kingdom specific heuristic models derived from more than 200 fungal genomes. The heuristic models also include G+C dependent sub-models for acceptor, donor and branch point sites. A subset of parameters, such as exon length distribution, was modeled with respect to bins with different intron density: low, medium and high.

For ab initio gene prediction in draft fungal genomes we have developed a group of fungi specific self-training methods. The constantly updated self- training tool GeneMark-ES has been used in a number of fungal genome sequencing projects since 2007. It remains to be a launch pad for integrated tools GeneMark-ET and GeneMark-EP that allow using mapped RNA-Seq reads and mapped homologous proteins, respectively, in training and prediction.

W218: Cool Season Legumes

Characterization of Photoperiod Response Genes in Chickpea

Stephen Ridge[1], Amit Deokar[1], Robyn Lee[2], Ketema Daba[1], Richard Macknight[2] and Bunyamin Taran[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)University of Otago, Dunedin, New Zealand

In climates that experience short growing seasons due to drought, heat or end-of-season frost, early flowering time is a highly desirable trait for chickpeas. In this study, we have mapped, sequenced and characterized what is likely to be *Efl-1*, an ortholog of the Arabidopsis gene *ELF3* that confers early flowering and photoperiod insensitivity. In a RIL population derived from a cross between CDC Frontier and ICCV 96029, this gene was mapped to the site of a QTL on Ca5 that explained approximately 60% of flowering time variation under SD. Sequencing in ICCV 96029 revealed an 11-bp deletion in the first exon that was predicted to result in a premature stop codon. The effect of this mutation was tested by transgenic complementation in the Arabidopsis *elf3-1* mutant, with the Frontier form of *ELF3* complementing *elf3-1*, while the 96029 form had relatively little effect on flowering time. Analysis of circadian clock function failed to show any clear loss of rhythm in the expression of clock genes in ICCV 96029 grown under continuous light, possibly suggesting an alternative mode of action for this gene in chickpea. This polymorphism does not appear to be widely-distributed within global chickpea germplasm, indicating that this mutation arose relatively recently. The implications of our findings for the development of early flowering chickpeas will be discussed.

W219: Cool Season Legumes

Molecular And Genetic Analyses Of Seed And Flower Color Variation In Chickpea

R. Varma Penmetsa, University of California at Davis, Davis, CA

TBD

W220: Cool Season Legumes

Development of Novel Genomic Resources in Pea (*Pisum sativum* L.) Revolutionizes Applications in Selection

Nadim Tayeh[1], Anthony Klein[1], Christelle Aluome[2], Matthieu Falque[3], Françoise Jacquin[1], Aurélie Chauveau[2], Aurélie Bérard[2], Hervé Houtin[1], Céline Rond[1], Marianne Chabert-Martinello[1], Jonathan Kreplak[1], Dominique Brunel[2], Marie-Christine Le Paslier[2], Grégoire Aubert[1] and Judith Burstin[1], (1)INRA, UMR1347 Agroécologie, Dijon, France, (2)INRA, US1279 Etude du Polymorphisme des Génomes Végétaux, CEA-IG / Centre National de Génotypage, Evry, France, (3)INRA, UMR320/UMR8120 Génétique Quantitative et Evolution - Le Moulon, Gif sur Yvette, France

Pea (*Pisum sativum* L.) is an important food and feed crop and a valuable component of low-input farming systems. Pea breeding has achieved great success since the time of Mendel's experiments in the mid-1800s. However, several traits still require significant improvement for better yield stability in a larger growing area. Key breeding objectives in pea include improving biotic and abiotic stress resistance and enhancing yield components and seed quality. Pea has recently benefited from next-generation sequencing and high-throughput genotyping technologies. The GenoPea 13.2K SNP Array, powered by Illumina Infinium® II technology, was developed. Twelve recombinant inbred line populations were genotyped using this SNP array and 12,802 transcript-derived SNP markers could be placed on a 15,079-marker high density, high resolution consensus map. This allowed (1) the localization of a large set of genes, (2) the identification of ohnolog-rich regions within the pea genome, and (3) the establishment of dense syntenic networks with sequenced legume genomes. The genomic prediction of important agronomical traits was evaluated in a collection of 339 genetic resource accessions (CRB339) genotyped using the GenoPea 13.2K SNP Array. Prediction quality was high even in cross-population experiments where the CRB339 collection was used as a training set and 9 recombinant inbred lines populations as a test set. The significant development and deployment of genomic resources is expected to enhance breeding by paving the way for genome-wide association studies and genomic selection approaches and also strengthen pea as a model for genetics and physiology.

This work was supported by the French National Institute for Agricultural Research (INRA) and funded by the French National Research Agency (Project ANR-09-GENM-026 "GENOPEA" and Project Investissements d'Avenir ANR-11-BTBR-0002 "PeaMUST").

W221: Cool Season Legumes

Assembly of the Pea Genome by Integration of High Throughput Sequencing (PacBio and Illumina) and Whole Genome Profiling (WGP™) Data

Mohammed-Amin Madoui[1], Karine Labadie[1], Léo d'Agata[1], Jean-Marc Aury[1], Jonathan Kreplak[2], Krishna Kishore Gali[3], Bunyamin Taran[3], Petr Capal[4], Jan Vrana[4], Caroline Belser[1], Marie-Christine Le Paslier[5], Rebecca McGee[6], David Edwards[7], Jacqueline Batley[8], Abdelhafid Bendahmane[9], Hélène Bergès[10], Grégoire Aubert[2], Valerie Barbe[1], Judith Lichtenzveig[11], Clarice J Coyne[12], Tom Warkentin[3], Jaroslav Dolezel[4], Patrick Wincker[1] and Judith Burstin[2], (1)CEA - Genoscope, Evry, France, (2)INRA, UMR1347 Agroécologie, Dijon, France, (3)University of Saskatchewan, Saskatoon, SK, Canada, (4)Institute of Experimental Botany, Olomouc, Czech Republic, (5)INRA, US1279 EPGV Etude du Polymorphisme des Génomes Végétaux, CEA-IG/Centre National de Génotypage, EVRY, France, (6)USDA-ARS, Pullman, WA, (7)University of Western Australia, Perth, Australia, (8)University of Queensland, Brisbane, Australia, (9)INRA, Evry, France, (10)INRA - CNRGV, Castanet Tolosan, France, (11)Curtin University, WA, Australia, (12)USDA-ARS Western Regional Plant Introduction Station, Pullman, WA

Pea (*Pisum sativum*) is one of the most cultivated grain legume crops in the world, however the pea crop encounters recurrent abiotic and biotic stresses that decrease its competitiveness. To increase our capacity to efficiently breed new cultivars and to better understand pea genome differentiation in different ecotypes, the International Consortium for Pea Genome Sequencing gathered their efforts to produce a high quality pea genome reference sequence. The pea genome being very large (4.5 Gb) and highly repetitive, the consortium strategy is based on the integration of different and complementary data such as whole genome sequencing data (short and long reads), physical map, optical map and genetic map data.

Whole genome shotgun DNA libraries of *Pisum sativum var*. Cameor with fragment sizes ranging from 160 bp to 17 kb were built and sequenced using the Illumina technology. A *de novo* assembly of the genome was done using a de Bruijn graph approach, contigs were scaffolded with the large insert reads and gap closed with PacBio reads. In parallel, whole genome profiling (WGP™) was used to create a physical map covering the entire genome. Based on the initial assembly, the WGP and the sequencing data, we improved the genome assembly continuity thanks to the MaGuS method (https://github.com/institut-de-genomique/MaGuS). The quality of the assembly was investigated by checking the collinearity with the genetic map and by using chromosome specific sequences obtained from chromosome sorting.

W222: Cool Season Legumes

Lentil 1.0 and Beyond

Kirstin Bett[1], Larissa Ramsay[1], Crystal Chan[1], Andrew G. Sharpe[2], Douglas R Cook[3], R. Varma Penmetsa[3], Peter Chang[4], Clare Coyne[5], Rebecca McGee[6], Dorrie Main[7], David Edwards[8], Sukhjiwan Kaur[9] and Albert Vandenberg[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)University of California at Davis, Davis, CA, (4)University of Southern California, Los Angeles, CA, (5)USDA ARS, Pullman, WA, (6)USDA-ARS, Pullman, WA, (7)Washington State University, Pullman, WA, (8)University of Western Australia, Perth, Australia, (9)Department of Economic Development, Jobs, Transport and Resources, Bundoora, Victoria, Australia

We have sequenced the lentil variety CDC Redberry using next generation DNA sequencing with paired-end, mate-pair and long read libraries over a wide range of sizes and different technologies. The assembly Lentil v1.0 consists of 7 pseudomolecules anchored through the use of six high-density genetic linkage maps, with the total assembled bases representing approximately half of the 4 Gb lentil genome. Over 49 thousand transcripts have been predicted on 15 thousand contigs through mapping of RNASeq and EST sequence data and are now available for BLAST and visualized on a JBrowse instance.

The completion of a draft genome encompassing about half of the lentil genome represents a powerful enabling platform for lentil biology, as we document in studies of genome sequence-assisted dissection of agronomic traits. The draft genome facilitates activities of a new project, AGILE, funded by the Saskatchewan Pulse Growers, Genome Canada, and partner agencies. The goals of AGILE are to understand genome evolution in the genus *Lens*, and the changes that have occurred during domestication and subsequent adaptation to different lentil growing regions of the world, and to accelerate breeding through the development of tools for molecular marker-assisted selection.

W223: Cool Season Legumes

Cool Season Food Legume Genome Database: An Up-to-Date Resource Enabling Genetics, Genomics and Breeding Research in Pea, Lentil, Faba Bean and Chickpea

Jodi L. Humann[1], Sook Jung[1], Ping Zheng[1], Chun-Huai Cheng[1], Taein Lee[1], Morgan Frank[1], Deah McGaughey[1], Kristin Scott[1], Jing Yu[1], Stephen P. Ficklin[1], Marwa N.M.E. Sanad[2], Heidi Hough[1], Clare Coyne[3], Rebecca McGee[1] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)National Research Center- Egypt, Pullman, WA, (3)USDA ARS, Pullman, WA

The new, mobile-friendly version of the Cool Season Food Legume Genome Database (CSFL, www.coolseasonfoodlegume.org) has been redesigned to allow for more efficient access to data, tools, and resources by users. The database has been updated with all current genetic maps, molecular markers, and QTL data in addition to the most current genome data for pea, lentil, chickpea, and faba bean. The new interface allows users to quickly search and retrieve data from the database. Quick access to popular tools which allow users to use BLAST for searches with current genome sequences and transcripts, view genomes in the JBrowse genome browser, make comparisons of genetic and physical map data with CMap, and view metabolic PlantCyc maps are also easily found from the website header. The ultimate goal of CSFL is to provide a single website where researchers can view/query/download all current genetics, genomics and breeding data for pea, lentil, chickpea and faba bean as well as have access to analysis tools that are useful for research. This project is supported by USDA NRSP10, the USA Dry Pea and Lentil Council, Northern Pulse Growers Association, USDA-ARS and Washington State University.

W224: Crop Genomics for Global Food Security

Barley Genomics and Food Security

Peter Langridge, Australian Centre of Plant Functional Genomics, Adelaide, Australia

W225: Crop Genomics for Global Food Security

Diversity Seek

Ruth Bastow, Global Plant Council, London, United Kingdom

Diversity Seek (DivSeek): An international partnership to harness the genetic potential of crop diversity

Peter Wenzl1, Ruth Bastow2, Daniele Manzella3, Wayne Powell4, Susan McCouch5
1 Global Crop Diversity Trust, Bonn, Germany
2 Global Plant Council, London, UK
3 Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture, Rome, Italy
4 CGIAR Consortium Office, Montpellier, France
5 Cornell University, Ithaca, USA

More than 1,700 genebanks globally conserve approximately seven million accessions of crop species and their wild relatives. In addition, farmers conserve and adapt germplasm to social, economic and ecological conditions across diverse farming systems. The genetic diversity encapsulated in all these materials underpins and drives crop improvement. Yet efforts to identify and mobilize beneficial genetic variation into breeding programs have been limited, compared to the size of this global resource. Game-changing 'omics' and 'big data' technologies now enable a more effective and broad-based approach to harnessing crop diversity. The recently launched Diversity Seek initiative (DivSeek; http://www.divseek.org) aims to capture this opportunity to accelerate the development of climate-ready, high-yielding and nutritious varieties for a growing global population. DivSeek provides a platform to generate synergies and add value to like-minded, but otherwise autonomous efforts by germplasm holders, breeders, geneticists, and database and computational experts to make crop diversity more readily accessible and usable. It is a community-driven initiative based on voluntary membership, which focuses on common challenges encountered by individual projects. Priority areas of work will include (i) the establishment of a 'cross-crop meeting platform' to share experiences and rapidly spread innovative research approaches and techniques, (ii) the development and advocacy of common data standards and interoperable data sets/repositories, (iii) a broadly accepted framework for 'rights management' that helps individual projects comply with data-sharing principles, and (iv) capacity-building efforts in these areas. We welcome colleagues and organizations interested in mining crop diversity for food security to join the growing group of the 60-some institutions that have come together to establish DivSeek.

Presenting Author
Ruth Bastow
Global Plant Council
Email: info@globalplantcouncil.org

W226: Crop Genomics for Global Food Security

Legume Genomics and Food Security

Henry T. Nguyen, University of Missouri, Columbia, MO

W227: Crop Genomics for Global Food Security

Potential and Challenges in Harnessing Public Genomic Research to Meet Global Food and Resource Needs

Shawn Kaeppler, Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI

Funding for crop genomic research has produced unprecedented information on plant genome structure and variation. Emerging initiatives seek resources to translate this information into improved cultivars with novel characteristics, and to provide enhanced predictive tools for growers via full season monitoring. Despite this exciting potential, the challenge and need has never been greater to find a path to market of novel research discoveries. Translation of basic information often faces the valley of death, due in large part to inexperience of scientists in the process of commercialization, relatively little entrepreneurial funding in the ag sector, and insufficient and non-uniform data to enable major companies to evaluate potential technologies. Furthermore, utilization of resources such as transgenic crops requires resources and training for appropriate stewardship and to meet regulatory requirements. Examples of the potential of genomics from our work in maize will be provided, and opportunities to meet potential challenges will be discussed.

W228: Crop Genomics for Global Food Security

The Genomic & Open-source Breeding Informatics Initiative

Kelly Robbins[1], Edward S. Buckler[2], Jean-Luc Jannink[1], Tobias Kretzschmar[3], Lukas Mueller[4], Yaw A. Nti-Addae[1], Michael S. Olsen[5], Mark E Sorrells[1], Qi Sun[6], Rajeev K Varshney[7] and Susan McCouch[1], (1)Cornell University, Ithaca, NY, (2)USDA-ARS-Cornell University, Ithaca, NY, (3)International Rice Research Institute, Los Banos, Philippines, (4)Boyce Thompson Institute for Plant Research, Ithaca, NY, (5)CIMMYT, Nairobi, Kenya, (6)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (7)ICRISAT, Hyderabad, India

In the last ten years, genotyping costs have dropped significantly, making feasible powerful new breeding approaches that can take advantage of the vast amounts of genomic data that have been generated in staple crops such as rice, wheat, maize, sorghum, and chickpea. The Genomic & Open-source Breeding Informatics Initiative (GOBII) is the first large-scale public-sector effort to enable systematic application of high-density genotypic information to the breeding of staple crops in the developing world. The project will develop and implement genomic data management systems to enhance the capacity of public-sector breeding programs to deliver increased rates of genetic gain in South Asia and Sub-Saharan Africa. The genomic data management systems will include databases, analysis pipelines, and decision support tools for plant breeders.

W229: Crop Genomics for Global Food Security

Crop Genomics in South America

Antonio Costa De Oliveira, Universidade Federal de Pelotas, Pelotas-RS, RS, Brazil
The need for increasing food production is a constant and challenging demand on crop breeding around the world. Major crops in South America have been benefiting from genomic approaches and production is constantly increasing. An x-ray vision on the potential expansion and limitations on major crops such as soybean, maize, wheat and rice is needed to draw our future agriculture. Global climate changes reflecting on novel biotic and abiotic stresses are the major challenges on crops for the upcoming years. Critical decisions on the use of GM crops have boosted soybean, maize and cotton crops, but advances in other crops are yet to be approved.

W230: CSSA: Translational Genomics
TBA
Mitch Tuinstra, Purdue University, West Lafayette, IN

W231: CSSA: Translational Genomics
Epigenetic Variation in Legume Crops
Scott A. Jackson, Robert Schmitz, Kyung Do Kim, Jonathan Corbi and Lexiang Ji, University of Georgia, Athens, GA
Epigenetic variation, that is non-sequence-based variation, has been of great interest in plant and animal research in the last few years. Little has been done to explore epigenetic variation in the context of plant improvement. Using soybean, we are exploring the level of epigenetic variation in soybean breeding germ plasm from across many US breeding programs. Using sodium bisulfite DNA sequencing, Methyl C sequencing, we have sequenced more than 100 soybean accessions including landmark cultivars from the past 80 years of improvement in the US, parents of the public NAM (nested association mapping) population, landraces and undomesticated soybean. Using these data, including RNA-seq and small RNAs, we determined differentially methylated regions (DMR) for all three methylation contexts (CG, CHG and CHH). These DMRs were then analyzed for their genomic context: exonic, UTRs, upstream, downstream, and intergenic or repeat (e.g. transposable elements). I will discuss these data in context of DMR variation (inherited and de novo) within and between varieties, landraces and wild soybean as well as their potential contribution to breeding/selection over the past 80 years and their contribution to variation.

W232: CSSA: Translational Genomics
The Effect of Artificial Selection on Phenotypic Plasticity: The Genotype by Environment Interaction Project in Maize
Natalia de Leon, University of Wisconsin-Madison, Madison, WI, Diego Jarquin, University of Nebraska-Lincoln, Lincoln, NE, M. Cinta Romay, Institute for Genomic Diversity, Cornell University, Ithaca, NY, Joseph Gage, University of Wisconsin, Madison, Madison, WI, Shawn Kaeppler, Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI, Edward S. Buckler, USDA-ARS-Cornell University, Ithaca, NY, Aaron J. Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN and G2F Consortium, G2F, Madison, WI
Remarkable productivity levels have been achieved in crop species through artificial selection and adaptation to modern agronomic practices around the world. The question remains as to whether this form of accelerated evolution has also enhanced the ability of such improved cultivars to consistently maintain high productivity across changing environmental conditions. A deeper understanding of the types of genetic architecture and modulation mechanisms controlling phenotypic plasticity, or genotype by environment (G X E) interaction, will enhance our ability to predict performance of improved crop varieties under ever-changing set of environmental conditions. This presentation will use the framework of the Genomes to Fields (G2F) G X E Maize project to assess the effect of selection on standing variation in genomic regions associated with G X E. In 2014, the G X E Maize project evaluated a collection of approximately 900 diverse maize genotypes across 23 sites in North America. Genotypes were evaluated for relevant phenological and agronomic characteristics. Climatic information was also collected in all locations. Genomic regions highly affected by selection explain a small percentage of the variability for plant height and yield G X E interaction compared to regions not differentially affected by selection. This presentation will focus on describing details related to these findings and a general description of the genetic architecture of G X E in this collection of materials.

W233: CSSA: Translational Genomics
Revitalizing Historic Soybean Mutant Resources using New Genomic Tools
Benjamin W. Campbell, Erin Gilbert, Amritpal Singh, Robert M. Stupar and Aaron J. Lorenz, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN
The recent genotyping of the USDA soybean germplasm collection has provide a wealth of data that is enabling new research goals and objectives. Starting in the 1960's, soybean researchers created a collection of near isogenic mutant lines, called the Soybean Isolines, by backcrossing one or more mutant alleles into the cultivars Clark, Harosoy, and Williams. While many of the mutant alleles in the Soybean Isoline collection are agronomically detrimental, some of the alleles have positive effects on yield, insect resistance, and disease resistance. The public release of the SoySNP50K genotyping data provided sufficient information to map the introgression intervals for many of the lines in the Soybean Isoline collection. For the 23 mutant alleles we have currently mapped, the introgression intervals range in size from 32kb to 5,541kb, with a median side of 606kb. The number of genes contained in these intervals range from 6 to 255, with a median gene number of 73. Proof-of-concept of this mapping method was performed on the trichome mutant T31, resulting in the cloning and validation of the causative sequence polymorphism. We are in the process of mapping and cloning additional genes using this method. This work provides a roadmap for furthering knowledge of soybean gene function and introgressing potentially valuable mutant alleles into elite germplasm.

W234: CSSA: Translational Genomics
Applying Genomic Resources to Restoration of an Iconic Forest Tree, the American Chestnut
Margaret Staton[1], Nate Henry[1], Jack Davitt[2], Nathaniel Cannon[3], C. Dana Nelson[4], Jared Westbrook[5], Jason Holliday[6], F. Hebard[5], L. Georgi[5], Albert G. Abbott[7] and John E. Carlson[3], (1)University of Tennessee, Knoxville, TN, (2)University of Tennessee, Knoxville, Knoxville, TN, (3)Pennsylvania State University, University Park, PA, (4)Southern Institute of Forest Genetics, US

Forest Service SRS, Saucier, MS, (5)The American Chestnut Foundation Meadowview Research Farms, Meadowview, VA, (6)Virginia Tech University, Blacksburg, VA, (7)Forest Health Research and Education Center, University of Kentucky, Lexington, KY

For over thirty years, scientists and private citizens have worked together to restore the American chestnut (*Castanea dentata*), a foundational forest species in eastern North America that was eliminated as a canopy tree in the first half of the twentieth century. Genomic resources have been established for American chestnut and Chinese chestnut (*C. mollissima*), a major source of genetic resistance to chestnut blight (*Cryphonectria parasitica*). Genomic resources, including a draft reference genome of C. mollissima, resequencing of additional chestnut species, genetic maps, and a physical map, are poised for immediate translation into the ongoing restoration effort. To identify candidate genes conferring blight resistance, we have performed resequencing of bacterial artificial chromosomes (BACs) spanning three blight resistance QTL regions. Comparison of the *C. mollissima* reference genome to ten other plant reference genomes has revealed major regions of conserved macro- and micro-synteny with other long-lived woody angiosperms, particularly peach, grape and poplar. To further the utilization and integration of genomic resources from chestnut and other hardwood tree species, a suite of online tools is now available, including a genome browser, a gene editor and a genome comparison tool at the Hardwood Genomics website (http://hardwoodgenomics.org). New initiatives utilizing these genomic resources are underway, including genotyping-by-sequencing (GBS) of thousands of trees in the American Chestnut Foundation backcross breeding program (http://www.acf.org/) and implementation of genomic selection.

W235: CSSA: Translational Genomics

Integrated and Translational Genomics for Analysis of Complex Traits in Crops

Ratan Chopra[1], Mark D. Burow[2], Gloria Burow[1], Charles Simpson[3], Jennifer Chagoya[2], John Burke[1] and Zhanguo Xin[4], (1)USDA-ARS, Lubbock, TX, (2)Texas A&M AgriLife Research, Lubbock, TX, (3)Texas AgriLife Research, Stephenville, TX, (4)USDA ARS, Lubbock, TX

We report here on integration of sequencing and genotype data from natural variation (by whole genome resequencing [wgs] or genotype by sequencing [gbs]), transcriptome (RNA-seq) and mutant analysis (also by wgs) with the goal of translating gems from these resources into useable DNA markers in the form of single nucleotide polymorphic (SNP) markers for crop improvement in peanut and sorghum. In peanut (*Arachis* sp), RNA-Seq was utilized to identify polymorphic SNPs in a cross between *A. duranensis* x *A. cardenasii*. Polymorphic SNPs were further used to develop a genetic map to identify QTLs for plant architectural traits including leaf measurements, main stem height, presence of main stem flowers, and seed weight. Validation of peanut architecture associated SNPs from this diploid population will be tested in populations of tetraploid peanuts. Meanwhile in sorghum, a genome wide association study for sorghum seedling stress was performed using GBS data and QTLs were identified for absorbance of flavonoids, shoot/root lengths, and shoot/root weights. SNPs associated with these traits in sorghum were screened in mutant germplasm and validated in a recombinant inbred population. To date there is an exponentially increasing amount of sequence and genotype data for major crops including peanut and sorghum, and integrated studies through a translational genomics approach and pipeline for identification of valuable SNP markers are vital for marker assisted breeding to become an integral part of germplasm enhancement.

W236: Cucurbit

Uncovering New Genetic Tools and Mechanisms in Mediating Carotenoid Accumulation *via* Investigation of Melon *CmOr* Gene

Li Li, USDA-ARS, Ithaca, NY and Yaakov (Kobi) Tadmor, Agricultural Research Organization, Newe Ya'ar Research Center, Ramat Yishay, Israel

Carotenoids are responsible for the orange flesh color found in various melon fruit. Our recent studies defined *CmOr* as the single gene that controls fruit flesh color and identified *CmOr* as the previously described *gf* locus in melon. *CmOr* was found to co-segregate with fruit flesh color and present with two haplotypes in a broad germplasm collection, one being associated with orange flesh and the second being associated with either white or green flesh. A single “golden” SNP that alters the evolutionarily highly-conserved arginine^{108} (Arg) to histidine (His) in CmOR governs increased b-carotene accumulation visualized as orange fruit flesh melon. Functional characterization of *CmOr* alleles of orange and non-orange fruits in transgenic Arabidopsis plants confirms the “golden” SNP role in controlling the inheritance of fruit color in melon. Moreover, we found that alteration of the nucleotides to change the conserved Arg into both histidine and alanine in *Or* homologs of other plant species also promote high levels of carotenoid accumulation, demonstrating the effectiveness of the “golden” SNP and its critical role in promoting carotenoid accumulation. The OR protein was discovered to possess dual functions in regulating both PSY for carotenoid biosynthesis and chromoplast differentiation to enhance storage sink strength. The metabolic and cellular processes associated with *CmOr* allelic variation and genes that participate in these processes were revealed *via* an F_3 family bulked segregant RNA-Seq (BSR-Seq) analysis. These findings broaden our fundamental understanding of mechanisms underlying carotenoid accumulation in plants, and provide effective genetic tools for developing agricultural products with improved b-carotene content.

W237: Cucurbit

Pending

Michael Mazourek, Cornell University, Ithaca, NY

W238: Cucurbit

Optimization of Melon Genetic Transformation and Genome Editing

Satoko Nonaka, University of Tsukuba, Tsukuba, Japan

Melon is the widely cultivated fruit vegetable species, and it has acquired a many kinds of fruit traits, including fruit shape, aroma, and ripening patterns during the domestication. Therefore melon provides an excellent material for studies of fruits traits, which are one of the most important issues in crop development. Melon is important in plant science, because melon served as an excellent model for investigating sex determination, ripening process, and vascular fluxes since xylem and phloem saps can be readily collected. For effective use of the attractive materials, functional genomics tools are required in melon. Therefore, for the last decade, many studies has been addressed the tools, such as expressed

sequence tag (EST) collections, and omics database, genetic linkage maps, DNA markers, bacterial artificial chromosome (BAC) libraries, a mutation library, a targeted induced local lesions in genomes (TILLING) platform for mutation screening, and transformation techniques. Although the transformation is a key technique for functional genomics research, the frequency is not enough for the practical use, due to the three problems such as tetraploidy, chimera, and escapes. Therefore the research efforts have been undertaken in order to establish the efficient transformation protocol in melon. Previous study showed the liquid culture system through somatic embryogenesis overcame these problems. In this study, to establish more sophisticated method, we improved the gene transfer method, that is, Agrobacterium strains, seed quality, the selection timing of embryogenic callus in two melon varieties. We have also tried to adopt this technique to the genome editing.

W239: Cucurbit

Genetic Dissection of the Climacteric Fruit Ripening in Melon

Pablo Ríos, Lara Pereira, Marta Pujol and **Jordi Garcia-Mas**, IRTA-CRAG, Barcelona, Spain

Melon provides an interesting model system to study fruit ripening, as both climacteric and non-climacteric varieties exist. The availability of the genome sequence and re-sequencing of melon varieties provide powerful tools to help characterizing genes involved in fruit ripening. A near isogenic line (NIL) population in the Piel de Sapo (PS, non-climacteric) background with introgressions from PI 161375 (SC, non-climacteric) was used to characterize two QTLs that provide climacteric ripening to PS. One QTL (*eth6.3*) has been cloned, and another QTL (*eth3.5*) has been mapped in a short genomic interval. A new RIL population obtained after crossing PS with the Védrantais (Ved) climacteric melon type revealed that climacteric ripening is a complex trait. The combination of these genetic resources with genomic tools will help dissecting the mechanisms that control climacteric ripening in melon.

W240: Cucurbit

Introduction to CucCAP - Developing Genomic Resources for the Cucurbit Community

Rebecca Grumet, Michigan State University, East Lansing, MI

The U.S. Cucurbit community has developed a USDA-SCRI funded cucurbit genomics project, CucCAP: Leveraging applied genomics to increase disease resistance in cucurbit crops. Primary objectives of the project include development of communal sequence and phenotype databases and bioinformatics tools for watermelon (*Citrullus lanatus*), melon (*Cucumis melo*), cucumber (*Cucumis sativus*) and squash (*Cucurbita pepo*). We will perform genotyping by sequencing (GBS) of approximately 1000 accessions from the PI collections of each of the four crops. Analyses of these data will be used to develop genome-informed core collections of 384 accessions for each species. Individual plants from the core collections will be self-pollinated and resequenced by GBS, providing a set of diverse lines and their associated sequence data, SNP datasets and genetic maps which will be available for future phenotypic and GWAS analysis of any traits of interest. Genomic and phenomic data and associated bioinformatic tools generated by the project will be publicly available through the International Cucurbit Genomics Initiative (ICuGI) website, along with workshops and webinars to facilitate use. The CucCAP project will focus on resistance to key diseases for each crop as identified by cucurbit producers. We will use a combination of GWAS and bi-parental populations to identify resistance associated QTL and initiate marker development. Breeding efforts for resistances to downy mildew, *Fusarium*, gummy stem blight, *Phytophthora capsici*, powdery mildew and several viruses will be performed in parallel with the genomic analyses and tool development.

W241: Cucurbit

Identification and Characterization of '*superfruiter*' - a New Melon Type

Ido Oz[1], Joseph Burger[1], Ari Schaffer[2,3], Efraim Lewinsohn[1], Vitaly Portnoy[1], Nurit Katzir[1] and **Yaakov (Kobi) Tadmor**[1], (1)Agricultural Research Organization, ARO, Ramat Yishay, Israel, (2)Agricultural Research Organization, ARO, Rishon LeZion, Israel, (3)Volcani Center, ARO, Bet Dagan, Israel

Phenotypic screening of an EMS mutation library identified a unique recessive mutant that produces five times more fruit than the wild-type. Fruit were small, third of the wildtype size, and seedless. We called this mutant *superfruiter* (*sf*). We analyzed the effect of *sf* in several genetic backgrounds and in all cases *sf* decreased fruit size, increased fruit number and on the average significantly increased yield. We identified *sf* plants and wild-type plants from several F_2 segregating populations. We extracted RNA from shoot meristems, stems, leaves, female flower and young fruit of these plants, bulked the RNA according to the phenotypes and conducted RNA-sequencing of these bulks. This approach identified the *sf* gene and suggested mechanism for its mode of action. With *sf* gene in our hands we can now develop custom made *sf* melon varieties.

W242: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

Introducing CyVerse

Parker Antin, University of Arizona, Tucson, AZ

The *iPlant Collaborative* was launched in 2008 with goal of designing, developing and employing the premiere U.S. cyberinfrastructure to support data driven inquiry. Initially charged with supporting the plant sciences community, with refunding of our supporting NSF grant in 2013 our mission expanded to encompass all areas of the life sciences. With tens of thousands of users working across the life sciences, we are renaming *iPlant* to ***CyVerse*** to reflect and acknowledge our expanded mission and user community. This talk will discuss the expanded vision of *CyVerse*, and also the central role that plant scientists will continue to play as part of our core mission to advance science through data driven discovery.

W243: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

Integrated Approach Towards Sequencing a Large and Complex Genome - iPlant Portal Facilitates Management of Big Data

Ming-Cheng Luo, Department of Plant Sciences, University of California, Davis, Davis, CA

Aegilops tauschii is one of the three diploid progenitors of common (hexaploid) wheat. Its genome is expected to be between 4 and 5 Gb and was estimated to contain nearly 90% repeated sequences. The aim of the National Science Foundation funded project “Sequencing the *Aegilops*

tauschii genome" (http://aegilops.wheat.ucdavis.edu/ATGSP/) is to generate a high-quality reference sequence for the genome of *Ae. tauschii* which ultimately serves as a reference for the wheat genome due to its close relationship with wheat. A bacterial artificial chromosome (BAC)-based physical map of accession AL8/78 was constructed. A minimum tiling path of 42,025 BAC clones was constructed and eight-clone pools across 3,578 BAC contigs and 2,000 singleton clones were sequenced with the Illumina MiSeq platform and scaffolds were assembled. We also generated whole genome shotgun sequences by Illumina HiSeq platform and the Pacific Biosciences platform. Pseudomolecules were constructed by integrating the assembled BAC pool sequence scaffolds, linkage maps, whole genome shotgun Illumina assembly, and PacBio sequences. The NSF-funded project involves several institutions nationally and internationally. The iPlant portal enables us to effectively manage the large volumes of diverse data generated within the project. The pseudomolecules are now available, we plan to use any of the tools provided at iPlant for further analysis.

W244: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

Evolnc: A Pipeline for Comparative Genomic and Transcriptomic Analyses of Long Non-Coding RNAs

Andrew D Nelson, School of Plant Sciences, University of Arizona, Tucson, AZ

Transcriptomic analyses from across eukaryotes have led to the conclusion that most of the genome is transcribed at some point in the developmental trajectory of an organism. One class of these transcripts is termed long noncoding RNAs (lncRNAs). Reported lncRNA repertoires in mammals vary, but are commonly in the thousands to tens of thousands of transcripts, accounting for ~90% of the genome. Recently, attention has focused on understanding the evolutionary dynamics of lncRNAs, particularly their conservation within genomes. To facilitate lncRNA discovery and comparative analyses at the genomic and transcriptomic level, we present Evolinc. Evolinc is a custom pipeline that identified long non-coding RNAs from transcriptome assembly files and then searches for homologs in other species. Using sequence similarity, Evolinc reconstructs families of homologous lncRNAs, aligns the constituent sequences, builds gene trees, and uses gene tree / species tree reconciliation to infer evolutionary processes. The novelty of this pipeline is that it allows the user to investigate factors affecting lincRNA diversity within a large number of species. This pipeline is scaleable, working on one to thousands of lncRNAs and can perform comparisons between both large and small genomes. For ease of use we have pre-packaged Evolinc into an instance available in the iPlant's Atmosphere cloud computing service, as well as in iPlant's Discovery Environment. Evolinc is useful not only for inferring mechanisms affecting lncRNA diversity, but also for identifying lncRNAs in non-model systems where genome data is available.

W245: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

The iPlant Agave Application Program Interface, High Performance Computing, and You (the Computationally Competent).

Eric Fritz-Waters, Department of Animal Science, Iowa State University, Ames, IA

iPlant has much to offer researchers, including those who are computationally competent and utilize the command line for their work. The graphical user interface (GUI) that is utilized through the web portal is laid upon an infrastructure that can be utilized and optimized through the iPlant Agave application program interface (API) in order to access the high performance computing made available through the Texas Advanced Computing Center and other resources. The next generation sequencing alignment and variant discovery pipeline utilized at Iowa State University by the Reecy Lab Group has been implemented in the iPlant Agave system and perl scripts have been created to streamline the process of taking advantage of having the pipeline available at iPlant. All of this can be done from the comfort of your own computer without having to directly log on to a server.

W246: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

How to do Big-Data Science – Data Management for Genomics

Jeremy D. DeBarry, University of Arizona, Tucson, AZ

Without appropriate training and resources, data and metadata management can be a significant barrier to investigation, collaboration, and publication. The iPlant Data Store is the backbone of iPlant's collection of data management tools, bioinformatics apps, and educational resources. Collectively, these form a free, scalable, and extensible, cyberinfrastructure to train and empower users for effective data management. The Data Store accommodates a broad range of skill levels, with web-based and command-line access, efficient data transfers, and metadata association and searching. Share individual files or datasets publicly, with specific collaborators, or manage data privately and securely. As part of a broader effort to support scientific project management and data publication, a pipeline for submissions to the NCBI SRA database has recently been added. The pipeline simplifies SRA submissions with a GUI, automated BioProject creation, iPlant-enabled data transfers, template-based metadata entry, and automatic checksum generation.

W247: Cyberinfrastructure for Life Science and Beyond – Scaling your science with iPlant

"We liked the shaver so much we bought the company": Federating iPlant to the UK and beyond!

Ryan Joynson and Anthony Hall, University of Liverpool, Liverpool, United Kingdom

After the success of iPlant collaborative project in the US, the BBSRC have funded the extension of the iPlant cyberinfrastructure across the Atlantic to the United Kingdom. A collaboration between The Universities of Liverpool, Warwick, Nottingham and The Genome Analysis Centre (TGAC) is working towards federation of the iPlant environment onto hardware based at TGAC that will be used to facilitate big data analysis for the UK life science community. The team at iPlantUK are also working to create and implement bioinformatics workflows within the GUI based discovery environment of iPlant, driven by the needs of the UK life science community. More specifically, at Liverpool we have been working to create RNA-seq workflows incorporating the Tuxedo suite of tools for RNA differential expression analysis along with re-creating and improving mapping by sequencing workflows for analysis of Barley (*Hordeum vulgare*) and Hexploid wheat (*Triticum aestivum)* using multiple strategies in order to make often difficult to reproduce work flows available in a user friendly way in a centralised location which will encourage and facilitate reproducible and robust research. Here we will describe our current projects and outline the future direction of iPlantUK.

W248: Degraded DNA and Paleogenomics

Population Genomics of Passenger Pigeons

Beth Shapiro[1], Richard Green[2], André Elias Rodrigues Soares[3], Ben Novak[3] and Russell Corbett-Detig[4], (1)Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, (2)University of California, Santa Cruz, Santa Cruz, CA, (3)UCSC/EEB, Santa Cruz, CA, (4)UC Berkeley, Department of Integrative Biology, Berkeley, CA, CA

The passenger pigeon was once the most abundant bird in North America, with flocks that, during the early and middle 19th century, were estimated to comprise up to three billion individuals. Less than 100 years later, however, passenger pigeons were extinct. To test the hypothesis that over-exploitation drove passenger pigeons extinct, and to explore the genomic consequences of rapid population decline, we sequenced and assembled 41 mitochondrial genomes and five high-quality nuclear genomes from passenger pigeons and, using new technologies for *de novo* genome assembly from Dovetail Genomics, a nuclear genome from band-tailed pigeons (the closest living relative of passenger pigeons). Our mitochondrial data indicate that passenger pigeon populations were growing exponentially at the time of their extinction and not in a period of decline. The nuclear genomic data reveal a striking, bimodal genomic pattern of genetic diversity with regions of extremely high and low diversity organized in distinct segments across the genome. We hypothesize that this pattern is a consequence of a long-term extremely large effective population size and consequent dominance of natural selection as an evolutionary force.

W249: Degraded DNA and Paleogenomics

Using Palaeogenomes to Calibrate the Evolutionary Histories of Ice Age Mammals

Love Dalen[1], Eleftheria Palkopoulou[2] and Pontus Skoglund[2], (1)Swedish Museum of Natural History, Stockholm, Sweden, (2)Department of Genetics, Harvard Medical School, Boston, MA

Genomic data is increasingly used to investigate the evolutionary histories of wild organisms, for example through inference of past changes in demography as well as divergence times among populations. However, the precision of such inferences depend on accurate and taxon-specific estimates of the genome-wide mutation rate. One way to address this is to analyze genomes sampled at different points in time. This approach offers a unique opportunity to directly estimate changes in genetic diversity through time, and thus to calibrate genome-wide mutation rates. The aim of this presentation is to provide empirical examples of palaeogenomic data sets that have been used to examine mammalian evolutionary histories, based on estimates of mutation rates calibrated using time-stamped Pleistocene and Holocene genomes.

W250: Degraded DNA and Paleogenomics

Ancient DNA Provides Novel Insights into the Origin of Dogs

Laurent Frantz, University of Oxford, Oxford, United Kingdom

Dogs were the first domestic animal and the only one domesticated prior to the advent of settled agriculture. Despite their importance in human history, little is clear about their geographic or temporal origins as multiple issues impair our ability to retrace their history. For example, the wild ancestor of domestic dogs (grey wolf) is widely distributed across New and Old world; thus, multiple populations may have been domesticated in different parts of the world. In addition, over the last millenniums, major human migrations could also have affected dog's ancestry, erasing ancient signal the genome of modern dogs. Nevertheless, by providing temporal information, ancient DNA can solve these issues (*e.g.* identify ancestry turnovers). Here we use multiple sources of ancient DNA such as the complete genome (>25x) of a 6,000 years old Neolithic dog from Newgrange (Ireland) as well as mtDNA fragments from over 50 ancient dogs from 10,000 BP to 3,000 BP. We combine this novel data set with modern nuclear and mtDNA genomes as well as SNP data from over 600 dogs. Our analyses illuminate multiple aspects the history of domestic dogs depicting a complex phenomenon with regional ancestry turnovers, and possibly multiple domestication origins.

W251: Degraded DNA and Paleogenomics

Ancient Whole Genomes, Migration and Insights into European Origins

Daniel G. Bradley, Trinity College Dublin, Dublin, Ireland

Petrous bones often yield endogenous DNA contents of 50% or more in ancient DNA analysis, most other skeletal elements give values of 1% or less. Targeting these bones makes the whole genome sequencing of ancient humans and other large mammals a feasible approach, even an imperative for rare and valuable samples. This presentation will discuss how recent analysis of ancient genomes from the Caucasus at the eastern edge of Europe has uncovered one of the major ancestral strands that coalesce in modern human populations of that continent. Also, ancient data from its western edge illustrate the far-reaching influence of prehistoric migrations carrying this genomic heritage from the east.

W252: Degraded DNA and Paleogenomics

A Paleogenomic Perspective on the Evolutionary History of Ice Age Equids

Peter D. Heintzman[1], Grant D. Zazula[2], Mathias Stiller[1], James A. Cahill[1] and Beth Shapiro[1], (1)Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA, (2)Government of Yukon, Department of Tourism and Culture, Whitehorse, YT, Canada

The evolution of equids – modern horses, zebras, and donkeys – is a classic example of macroevolution inferred from the fossil record. Although the relationships between living equid species are well understood, substantial disagreement remains between morphology and genetics regarding the placement and taxonomy of their recently extinct kin. Here, we applied a paleogenomic approach to determine the placement of an extinct group of poorly understood, late Pleistocene equids. Based on whole mitochondrial and low coverage nuclear genomes, we demonstrate the robust, novel phylogenetic placement of this taxon. We explore the implications of this placement on our understanding of recently extinct equid evolution and fossil equid systematics.

W253: Degraded DNA and Paleogenomics

The Nature and Constraints of Plant Evolution Under Domestication

Robin G. Allaby, University of Warwick, Warwick, United Kingdom

Our understanding of the evolution of domestication has changed radically in the past ten years from a relatively simplistic rapid origin scenario to a protracted complex process. Through the use of archaeogenomics we can observe genome evolution directly through time. We have been applying this to understand how domesticated crops have evolved and adapted to their environments. Through NGS approaches we have retrieved genomic, transcriptomic and epigenomic data that are helping us to build a picture of how crops have evolved. The use of both archaeological and historical samples have proved crucial, as well as trawling the bottom of the sea for sedaDNA to track the spread of crops. We have found that historic samples can provide a clearer signal of adaptation than we can gain from modern samples that have been the subject of recent movements. Together, our data are producing a picture of how plants adapted as agriculture spread, and tempered that spread. Furthermore, we see the proactive adaptation of plants overriding the artificial selection pressures applied by humans illustrating the co-evolutionary nature of our relationship with the plant world.

W254: Degraded DNA and Paleogenomics

Origins of a "Mixed Up Monkey": Phylogenomic Analyses of the Type Specimen of the Enigmatic Lemur *Mixocebus*

Ian Barnes, Natural History Museum, London, United Kingdom

Natural history collections present excellent opportunities to resolve evolutionary processes in rare and extinct taxa. While the DNA from museum specimens is typically heavily modified, the application of next generation sequencing approaches has significantly our ability to work with this genetic material, which is characteristically highly fragmented and chemically damaged. Here, we have investigated the origins of the lemur genus *Mixocebus*. The genus is currently composed of a single specimen held in Berlin Museum since 1873, and has been sometimes considered as a missing link – or perhaps a hybrid -between the lemur families Lepilemuridae and Cheirogalidae. Our objective for this project was therefore to assess whether *Mixocebus* can be classified as a unique, extinct species; a morphologically distinct, but previously described species; or a cross-genus hybrid. Here, we discuss the results of our molecular work, focusing on the difficulties of resolving the phylogenetic position of a taxa for which relatively little comparative data is available, DNA preservation is poor, and the volume of starting material is extremely small.

W255: Degraded DNA and Paleogenomics

Ancient Tuberculosis in the Americas

Anne Stone[1], Tanvi Honap[1], Ashild Vagene[2], Alexander Herbig[2], Michael Rosenberg[1], Kirsten Bos[2], Jane Buikstra[1] and Johannes Krause[2], (1)Arizona State University, Tempe, AZ, (2)Max Planck Institute for the Science of Human History, Jena, Germany

Mycobacterial diseases, such as tuberculosis (TB) have long impacted humans, as well as our domesticated animals in the Old World. However, the origin of TB in the New World has been a long-standing topic of debate. Ample skeletal evidence for TB is present in pre-contact South and North America, but today New World TB is caused by *Mycobacterium tuberculosis* complex (MTBC) strains of European origin, suggesting that pre-existing MTBC strains were replaced following European contact. Previous research from our group led to the recovery of MTBC genomes from three 1000-year old skeletal TB cases from coastal Peru. The ancient Peruvian strains are distinct from any known human-adapted TB strain and are most closely related to strains adapted to sea mammals (specifically Southern Hemisphere pinnipeds). However, it remains unknown whether such pinniped-derived MTBC strains spread to inland parts of South America as well as North America by human-to-human transmission or whether different strains spread into North America via another route. The present work focuses on skeletal TB cases from pre-contact, protohistoric, and historic sites from Alaska, the Midwestern United States, and Colombia. DNA was extracted using a silica-based method and tested for presence of MTBC DNA using quantitative PCR assays. Fifteen DNA extracts tested positive for a region of the *rpoB* gene specific to the MTBC and the IS6110 and IS1081 repeat elements. In-solution target enrichment and sequencing of MTBC-specific genes showed good coverage as well as authenticated the presence of ancient DNA for these samples. Some of the samples that tested positive were enriched for the entire MTBC genome using an array-capture method and then sequenced. Analyses of these genomic data are currently ongoing.
This work was supported by funding from the National Science Foundation (BCS-1063939 and BCS- 1515163) and the Wenner-Gren Foundation.

W256: Development and Application of Transgenic Technology in Agriculture

Utilization of CRISPR/Cas9 to Study Meristem Growth and Shoot Architecture in the *Solanaceae*

Joyce Van Eck, Boyce Thompson Institute for Plant Research, Ithaca, NY

Tomato has become a valuable model for biological, genetic, and functional genomics studies. Its adoption as a model resulted from readily available resources such as mutant collections, a high quality genome sequence, mapping populations, and efficient transformation methodology. These resources provided the ideal platform for testing the feasibility of the genome editing technology, CRISPR/Cas9. We started with a proof-of-concept experiment to target a gene (*ARGONAUTE7; SlAGO7*) that when function was lost would result in an easily recognizable phenotype (needle-like leaves; wiry) evident as early as the plant regeneration stage following *Agrobacterium*-mediated transformation of cotyledon explants. The CRISPR construct contained two single guide RNAs (sgRNAs) to produce large deletions to ensure disruption of gene expression. Twenty-nine independent transgenic lines were recovered and 48% exhibited the expected wiry phenotype, which indicated the successful disruption of *SlAGO7*. DNA sequencing of PCR amplicons revealed the CRISPR lines were comprised of a range of mutations including homozygous, biallelic or chimeric small insertions and deletions (indels). The indels were present at various locations near both sgRNA targets. Following this success, we have designed more than 40 CRISPR constructs to target genes that were shown through RNAseq analysis to possibly play a role in flowering, stem maturation, and inflorescence branching. The CRISPR/Cas9 technology has been a powerful tool for our reverse genetics approach to elucidate the roles of these genes. Genes found to be of interest in tomato will also be targeted in other *Solanaceae* members to determine cross species effects on meristem development and shoot architecture.

W257: Development and Application of Transgenic Technology in Agriculture

RNAi Mediated Silencing of Endogenous Wheat Genes eIF4(iso)E-2 and eIF4G Induces Resistance to Potyviruses Wheat Streak Mosaic Virus and Triticum Mosaic Virus

Jessica L. Rupp, Montana State University, Bozeman, MT 59717-3150, MT, Luisa Cruz, Pathway Biologic, Plant City, FL, John Fellers, USDA, Manhattan, KS and Harold N. Trick, Kansas State University, Manhattan, KS

Wheat streak mosaic virus (WSMV) and *Triticum mosaic virus* (TriMV) are two viruses affecting wheat in the Great Plains of the United States. Current disease management strategies incorporate deployment of resistant varieties, mite vector control and various cultural practices; however, it is not fully effective. Both of these viruses belong to the family *Potyviridae* and use host eukaryotic initiation factors in order to facilitate replication of their genomes. We evaluated the use of RNAi to silence eIF4(iso)E2 and eIF4G to interrupt this process in order induce resistance to these wheat viruses. RNAi expression vectors were independently created from the sequences of the wheat genes eIF4E(iso)2 and eIF4G. Immature embryos of the wheat cultivar 'Bobwhite' were independently co-transformed by biolistic particle delivery system with RNAi expression vectors and pAHC20, which contains the *bar* gene for glufosinate selection. All progeny have undergone PCR and RT-PCR analysis. Progeny were mechanically inoculated with the viruses. A consistent stable resistance response was demonstrated in three transgenic lines of eIF4(iso)E2 construct and four transgenic lines of eIF4G, each derived by single seed descent. T_6 progeny were co-infected with WSMV and TriMV continue to be resistant. Traditional crosses have been performed with the winter wheat 'Overley.' Effectiveness of the RNAi construct has been evaluated using Real-time PCR. Results show up to 18-fold reduction in viral titer in the transgenic lines, the F1 cross and the BC1F1 in compared to control plants. This research provides evidence that a single transgene can provide resistance to multiple viruses having great potential benefits to breeders and producers.

W258: Development and Application of Transgenic Technology in Agriculture

Expression of Heat-Stable Starch Synthase Genes Increase Yield Potential of Heat Stressed Wheat

Harold N. Trick[1], Bin Tian[1], Shymal K. Talukder[2], Hyeonju Lee[1] and Allan K. Fritz[1], (1)Kansas State University, Manhattan, KS, (2)The Samuels Robert Noble Foundation, Ardmore, OK

Several global climate models predict that global temperatures will continue to rise this century, which is a major concern for wheat production worldwide. Wheat is a temperate cereal with an optimum growth temperature range of 15-22°C. During the grain filling stage of development, for every 1°C rise above the optimum temperature, yield is reduced ~3-6%. This heat stress reduces the kernal weight of wheat seeds by decreasing starch production. The soluble starch synthase (SSS), a key heat-labile enzyme, plays an important role in regulating the conversion of sucrose to starch in the wheat endosperm. Expression of putative thermostable SSS in genetically engineered wheat could increase the productivity under heat stress. We demonstrated that the grain weight of transgenic wheat was significantly improved under high temperature conditions by independently expressing three different thermostable SSS genes under the regulatory control of either the maize ubiquitin or wheat high molecular weight glutenin promoter. Under optimum growing conditions all agronomic traits evaluated (seed size, seed number, tiller number, and physiological maturity) had no significant variations. However, under heat-stress conditions the concentration of soluble starch in transgenic lines was significantly higher compared to the controls. Transgenic lines demonstrated up to a 34% increase in thousand seed weight compared to non-transgenic controls under heat stress conditions ranging from 31/24°C (d/n) to 34/28°C (d/n) during the grain filling period. Further analysis of transgenic wheat plants is in the process in hope to provide a novel strategy for improving heat tolerance for cool season crop plants.

W259: Development and Application of Transgenic Technology in Agriculture

Directing the Expression of Thermostable Glycohydrolases *in planta* and Designer Glycopeptide Engineering Technology for Sustainable Sugar Beet Post-Harvest Processing

Jianfeng (Jay) Xu, Brett J. Savary, Jose C. Tovar and Ningning Zhang, Arkansas State University, Jonesboro, AR

Sugar beets are targeted for expanded industrial sugar production beyond traditional growing regions to meet national needs for advanced biofuels, renewable chemical feedstocks, and for conversion to value-added biobased products. We are investigating biochemical and molecular technologies for controlled sugar beet root cell wall modification and hydrolysis to provide economic and environmental benefits for sustainable beet-biomass processing. We are assembling a toolbox with thermostable glycohydrolases, including a *Citrus* thermally-tolerant pectin methylesterase, a *Geobacillus* thermostable endo-arabinanase, and the *Acidothermus cellulolyticus* E1 endoglucanase, to evaluate effective action on beet pulp processing and demonstrate bioproduction systems. Our goal is to establish their suitability to rationally manipulate the structural and functional properties of structural polysaccharides *in situ* with their direct expression in sugar beet tap-roots.

Our experimental results showed that pectin methylesterase action dramatically reduces water binding in beet pulp through improved calcium binding. Recombinant arabinanase produced by yeast was able to generate functional oligosaccharides from sugar beet pulp, specifically, feruloylated arabino-oligosaccharides (which may promote healthful colon functioning through prebiotic, anti-inflammatory, and mucosal immuno-modulatory activities). We also leveraged hydroxyproline-*O*-glycosylation – a process unique to plant cell wall glycoproteins – as an innovative technology for *de novo* design and engineering of novel glycopeptides that can function as a molecular carrier for *in planta*-expressed cell wall-modifying enzymes (e.g., endoarabinanase), efficiently targeting enzymes to the cell wall matrix and conferring improved stability to the enzymes. Our results demonstrate proof-of-principle for ultimate recombinant enzyme expression in sugar beet roots, which will provide a novel output trait for biotech beets.

W260: Development and Application of Transgenic Technology in Agriculture

Mobile Small Interference RNAs and Grafting of Genetically Engineered Cherry Plants

Guo-Qing Song, Michigan State University, East Lansing, MI

Grafting is a well-established agricultural practice, and it now has implications for the commercialization of transgenic plants for non-transgenic products because in trans-grafted plants, only one part (scion or rootstock) is transgenic with the other part untransformed. In our recent studies, we transformed a hairpin RNA (hpRNA) vector to a major sweet cherry rootstock cv. Gisela 6 in order to silence *Prunus necrotic ringspot virus* (PNRSV), which is a major pollen-disseminated Ilarvirus. All transgenic rootstocks showed accumulation of hpRNA-derived small interfering RNAs (siRNAs) and high resistance to the PNRSV. Subsequently, we performed grafting studies to investigate whether PNRSV-resistant transgenic rootstocks developed through siRNA-mediated gene silencing can enhance virus resistance of non-transgenic scions. We found transported (rootstock-to-scion) siRNAs in a non-transgenic scion of sweet cherry grafted on a transgenic rootstock. More importantly,

inoculation of non-transgenic scions with PNRSV revealed that the transferred siRNAs enhanced PNRSV resistance in the scions grafted on the transgenic rootstocks. Low amounts of transferred hpRNA siRNAs in scions, compared to those detected in PNRSV-infected but symptomless cherry plants using non-transgenic rootstocks, implied little concern of these siRNAs for food safety. These findings provide the basis for 'using transgenic rootstocks to produce non-transgenic products of scions in rootstock-scion grafted plants', while minimizing concerns about food and environmental safety.

W261: Domestication Genomics

Olive Domestication and Diversification in the Mediterranean Basin

Concepcion M. Diez, University of Cordoba, Cordoba, Spain and **Brandon S. Gaut**, University of California, Irvine, Irvine, CA

Among domesticated plants, a major distinction is the difference between annual and perennial life cycles. The domestication of perennials is expected to follow different processes than annuals, with distinct genetic outcomes. Important issues for understanding the process of domestication – such as genetic bottlenecks and introgression as a source of local adaptation - have been studied nominally in major annual crops but even less extensively in perennials. Here I will focus on one perennial, olive (*Olea europaea* ssp. *europaea*) and a study of its genetic diversity within the Mediterranean Basin. The study provides insights into the complex demographic history of olive and reveals the potential for multiple domestication events and/or local introgression.

W262: Domestication Genomics

The Role of DNA Methylation in the Domestication of Tomato

Catarina Lira, Instituto de Pesquisas, Rio de Janeiro, Brazil, Carolina Voloch, Departamento de Genética, Rio de Janeiro, Brazil and **Amy Litt**, University of California Riverside, Riverside, CA

Epigenetic modifications contribute to phenotypic variation through their effect on phenomena such as genome stability and patterns of gene expression. Large scale epigenome studies have been carried out on both model and non-model plant species but little is known about the role of epigenetics in the phenotypic changes that occur during the domestication process. We used bisulfite sequencing to compare DNA methylation patterns in two accessions of cultivated tomato (*Solanum lycopersicum*) and three accessions of its putative wild ancestor (*S. pimpinellifolium*). All epigenomes were aligned with the reference genome of *S. lycopersicum* 2.50 from the Sol Genomics database using Bismark. Mapping efficiency was high for all libraries. Patterns of cytosine methylation were similar in all accession: 52%, 45% and 3% for CpG, CHG and CHH respectively. Cytosines in the CpG and CHG contexts were highly methylated (95% and 88% respectively). Overall cytosine methylation patterns were not correlated with repeat or centromere regions. Despite similar overall methylation rates, several genes were shown to be differentially methylated between the domesticated and wild accessions, including genes implicated in transcription regulation, flower development, and fruit development and ripening, as well as genes involved in metabolism, histone modification, and other cellular processes. To determine if the differences in methylation are correlated with differences in expression, we used quantitative RT-PCR to compare the expression of these genes in wild and cultivated accessions. Results will be presented and the potential role of these genes in domestication traits will be discussed.

W263: Domestication Genomics

Postive and Relaxed Selection on Chickpea during Domestication and Post-Domestication Diversification

Eric von Wettberg[1], Peter Chang[2], Vasantika Singh[2], Matilde Cordeiro[2], Alex Greenspan[3], Betsy Alford[4], Noelia Carrasquilla[3], Emmanuel Dacosta-Calheiros[1], Emily Warschefsky[1], Raiz Rouf Mir[5], Bekir Bukun[6], Abdullah Kahraman[7], Abdulkadir AydoÄŸan[8], Jens D. Berger[9], Sergey V Nuzhdin[2], R Varma Penmetsa[10] and Douglas R Cook[10], (1)Florida International University, Miami, FL, (2)University of Southern California, Los Angeles, CA, (3)University of California, Davis, CA, (4)University of California Davis, Davis, CA, (5)SKUAST JAMMU, JAMMU, India, (6)Dicle University, Diyarbakir, Turkey, (7)Harran University, Urfa, AZ, Turkey, (8)Turkish Ministry of Agriculture, Ankara, Turkey, (9)Plant Industry, CSIRO, Floreat, WA, Australia, (10)University of California at Davis, Davis, CA

Like many other annual crops, chickpea (*Cicer arietinum*) underwent a substantial population bottleneck when domesticated from its wild progenitor (*Cicer reticulatum*) approximately 10 thousand years ago. During this bottleneck several key domestication traits such as indehiscence were positively selected by early farmers. However, the small effective population sizes of early agriculture and the ecological transition to the protected habitat of a cultivated field likely relaxed selection on much of the cultivated chickpea genome. We document both positive and relaxed selection on the cultivated chickpea genome, utilizing a recently assembled expanded collection of the wild progenitor of chickpea as well as its sister taxa *Cicer echinospermum.* A thorough survey of the source populations for the wild relatives as well as their habitats allows inference into the levels of standing variation available to early farmers, and environmental factors that shifted with the advent of agriculture in the fertile crescent. Our work shows the need for in-depth collections of crop wild relatives and landraces shielded from modern breeding to make informed inferences about domestication.

W264: Domestication Genomics

Sequence and Expression Divergence during the Domestication of Eggplant

Anna M. L. Page[1], Rachel S. Meyer[2], Jane Gibson[1] and **Mark A. Chapman**[1], (1)University of Southampton, Southampton, United Kingdom, (2)New York University, New York, NY

Given the need to feed an increasing population, it is vital to identify the alleles, genes, cultivars and crops which confer yield and stress-tolerance characteristics. Discerning the location and timing of crop domestication allows wild progenitors to be located, and helps identify the genetic changes that confer the traits that humans selected for thousands of years ago. Further, the study of domestication arguably serves as a contemporary and tractable model for evolutionary processes including the effect of selection on the genome and whether the targets of selection are coding or regulatory.

Here we investigate the domestication of eggplant using RNAseq and Genotyping-by-Sequencing. Comparisons of geographically diverse samples of the progenitor species and eggplant reveals an Indian origin, and does not support the previous hypothesis of a second domestication

in China. Instead, individuals of the 'progenitor' species in China appear to be feral escapes from domestication, with evidence for hybridisation. Nucleotide diversity in eggplant was about half of that in the progenitor, indicating a moderate genome-wide domestication bottleneck. The RNAseq data were used to investigate the targets of selection during domestication, both at the gene expression level and the sequence level. Analysis reveals over 80% of the ca. 2,000 differentially expressed (DE) loci were expressed at a lower level in the domesticated species with many genes putatively silenced. GO terms related to metabolism, development and response to stress were over-represented in the DE loci. The sequence-based tests for selection revealed similar GO terms were over-represented. Selective sweep analyses are underway to confirm the targets of selection.

W265: Domestication Genomics

Genomics Reveals the Past and Informs the Future of Apple Improvement

Zoë Migicovsky[1], Kyle M. Gardner[1], Daniel Money[1], Jason Sawler[2], Joshua S. Bloom[3], Peter Moffett[4], C. Thomas Chao[5], Christopher M. Richards[5], Heidi Schwaninger[5], Gennaro Fazio[5], Gan-Yuan Zhong[5] and Sean Myles[1], (1)Dalhousie University, Truro, NS, Canada, (2)Anandia Labs, Vancouver, BC, Canada, (3)UCLA, Los Angeles, CA, (4)Université de Sherbrooke, Sherbrooke, QC, Canada, (5)USDA-ARS, Geneva, NY

Apple (*Malus domestica*) is one of the world's most valuable fruit crops and a promising candidate for marker-assisted selection (MAS) due to its lengthy juvenile phase. To examine the genetic structure of apples and discover genotype-phenotype associations, we generated over 8,000 SNP genotypes from 689 *M. domestica* accessions from the USDA germplasm collection using genotyping-by-sequencing (GBS). The primary axes of genetic differentiation in apples separate Old World from New World cultivars, and early from late harvested cultivars. We confirm that apples used for making alcoholic cider derive part of their ancestry from *M. sylvestris*, the wild European crab apple. We also find a complex pedigree network among cultivars that suggests extensive use of elite cultivars in recent apple breeding. The extremely rapid linkage disequilibrium (LD) decay in apple suggests that whole genome sequencing will be required for well-powered genome wide association (GWA). Despite this, we conducted GWA using over 24,000 phenotypic data points from the USDA-GRIN database and found significant hits for skin colour and fruit firmness, both of which show the genetic signature of improvement through breeding. Most notably, we found an amino acid substitution in the transcription factor NAC18.1 that is a strong functional candidate for fruit firmness. We demonstrate that GWA in the USDA apple germplasm collection results in extremely high resolution mapping of causal variants, which holds great potential for continued improvement of apples through MAS.

W266: Domestication Genomics

More Than Skin Deep: Limb Identity and the Origins of Feathered Feet in Domestic Pigeons

Michael D. Shapiro, University of Utah, Salt Lake City, UT

The rock pigeon (*Columba livia*) is among the most phenotypically diverse and widely distributed avian species in the world. We are combining traditional genetics, whole-genome sequencing, and developmental analyses to understand the evolution of diversity among the hundreds of breeds within this species. In doing so, we are addressing fundamental questions about the genetic architecture of phenotypic change, including: How many genetic changes are required to yield pronounced anatomical changes? What are the identities of these key genes? What types of mutations underlie phenotypic variation (e.g., coding versus regulatory)? How do different loci interact to produce complex traits? Do similar traits in different breeds evolve via similar or different genetic mechanisms? In particular, our combined approach is revealing a spectrum of mutations that control the orientation, color, and placement of plumage within and among breeds. In one dramatic example, genome sequencing across breeds revealed variation at two key limb genes that was associated with the replacement of scales by feathers on the feet. Laboratory crosses and developmental studies showed that cis-regulatory changes in these genes interact to produce complex epidermal appendage variation. These epidermal changes, combined with alterations in muscle patterning and skeletal morphology, represent a partial homeotic shift in the hindlimb to a more forelimb-like identity. The genomics era is driving new discoveries about the molecular basis of diversity in non-traditional model organisms. The domestic pigeon is a promising model with which to explore the genetic architecture of derived, constructive phenotypes in a bird that is amenable to genetic, genomic, and developmental investigation.

W267: Ecological Genomics

Variation and Selection of Genes Controlling Ecologically Important Traits in Nature

Julius P. Mojica and Thomas Mitchell-Olds, Duke University, Dept. of Biology, Durham, NC

Amidst the prevalence of studies on balancing selection or deleterious polymorphism, the relative importance of these evolutionary processes for phenotypic variation is unclear. To understand the evolutionary forces that influence complex trait variation in a wild relative of Arabidopsis, we cloned an ecologically important QTL in natural populations of *Boechera stricta* and measured the fitness of alleles in the populations where they evolved. Field measurements of selection indicate that this polymorphism is influenced by spatially heterogeneous natural selection. Next, we examined the relationship between flux and polymorphism in this pathway, showing that flux control is focused in the first enzymatic step, encoded by a gene experiencing selective diversification in several related species. Finally, to identify the genes and alleles responsible for ecologically important trait variation in nature, we are conducting a Genome-Wide Association Study (GWAS) on a panel of reference genotypes from 350 populations across the species range. For these accessions we assess the genetic variation from genome-wide resequencing and genotyping by sequencing, and associate these with trait variation for defensive chemistry, resistance to multiple herbivores, flowering time, and complex traits evaluated in the lab and the field. Implications for ecological genomics will be discussed.

W268: Ecological Genomics

Potential Paths of Migration and Genetic Basis of Herbicide Resistance in a Noxious Agricultural Weed

Regina S. Baucom, Diego Alvarado Serrano and Megan van Etten, University of Michigan, Ann Arbor, MI

Herbicide resistance is an unfortunate consequence of relying on a single method of weedy plant control. The evolution of resistance in natural plant populations will depend on various factors, such as the presence of genetic variation and the potential for gene flow between populations. Here we examine migration potential between populations of the common morning glory, *Ipomoea purpurea*, that show variable levels of

resistance to the herbicide glyphosate, using a combined genetic analysis of 15 SSR loci and 8210 SNPs. Our analyses show strikingly different patterns of population connectivity in *I. purpurea*. While the SSR loci supported an effectively panmictic population, the SNP loci provided evidence of highly structured populations, with localities isolated primarily by geographic distance and landscape cover between sampled localities. Such different pictures of the population dynamic of this species could be attributed to the different temporal scope and coalescent history of both sets of markers, yet further analyses are needed to address this hypothesis. Furthermore, we scanned the genome for signals of adaptation to glyphosate using BAYESCAN to identify loci that were more differentiated between resistant and susceptible populations than expected. Our results yield novel insight into the genetic basis of resistance in this species as well as the potential avenues of migration between populations.

W269: Ecological Genomics

Defining the Relationships Between Arabidopsis Innate Immunity Receptors and Phytopathogenic Immune Elicitors with Phylogenomics

Adam Mott, Shalabh Thakur, Pauline Wang, Darrell Desveaux and David Guttman, University of Toronto, Toronto, ON, Canada

The plant innate immune response begins with the binding of conserved microbial features called microbe-associated molecular patterns (MAMPs) to specific plant receptors. While the importance of these events has long been understood, traditional approaches to identify the involved molecules remain problematic. The recent profusion of genome sequences from bacterial phytopathogens has allowed for the development of bioinformatics approaches to MAMP identification. Using a comparative genomics approach, we have identified new peptide MAMPs from the phytopathogenic bacterium *Pseudomonas syringae* and utilized a reverse genetic screen to identify *Arabidopsis thaliana* immune receptors required for their recognition. Using a novel high-throughput assay we have tested the activity of seven bacterial MAMPs on 187 immune receptor T-DNA insertion knockouts. This primary screen has allowed us to identify clades of genes that play a role in increasing or decreasing MAMP responses globally, as well as lines unable to respond to a single MAMP, suggesting a role in MAMP binding. We focused on the interaction between the xup25 MAMP from a bacterial xanthine/uracil permease and the Arabidopsis receptor xanthine/uracil permease sensing 1 (XPS1). We showed that xup25 treatment results in a classic immune response, including pathogenesis-related gene induction, callose deposition, and resistance to virulent bacteria, all in an XPS1-dependent manner, indicating that XPS1 is specifically required for xup25 recognition. The study demonstrates an efficient method to identify immune elicitors and the plant receptors responsible for their perception. Further exploration of these molecules will increase our understanding of plant-pathogen interactions and the basis for host specificity.

W270: Ecological Genomics

Symbiotic and Transcriptomic Dimensions of *Trifolium* Coexistence

Maren L. Friesen, Michigan State University, Lansing, MI

W271: Ecological Genomics

iMicrobe: Extending the iPlant Cyberinfrastructure for Metagenomic Analysis in Microbial Ecology

Bonnie Hurwitz, University of Arizona, Tucson, AZ

W272: Engineering NUE

Utilize the Genetic Potential: Breeding Progress Towards Nitrogen Use Efficiency in *Brassica napus* L

Andreas Stahl[1], Benjamin Wittkop[1] and Rod Snowdon[2], (1)Justus Liebig University, Giessen, Germany, (2)Department of Plant Breeding, Justus Liebig University, Giessen, Germany

Fertilization with nitrogen is one of the most agronomic important factors contributing to maintenance of the high seed yields needed to match food and non-food demand for a strongly increasing world population. However, significant quantities of fertilized N escape into groundwater, rivers and ocean or in the form of volatile losses into the atmosphere. Although nitrogen fertilization will remain essential to achieve high yield levels in future, production of agricultural commodities must simultaneously reduce negative impacts on ecosystems.

The recent allopolyploid oilseed rape (*Brassica napus* L.), the third-most important oil crop worldwide after perennial oil palm and soybean, has multiple uses in food and non-food products. Furthermore, oilseed rape cultivation is part of a sustainable crop rotation. Compared to other crops, however, oilseed rape releases a high nitrogen balance surplus requiring improvements in nitrogen use efficiency (NUE).

We assessed diversity for NUE at six locations across Germany. 30 elite inbred and hybrid varieties, covering 25 years of breeding progress, were grown under two distinct nitrogen fertilization levels (120 and 220 kg N/ha). Alongside seed yield and quality parameters we also conducted a biomass harvest at flowering. This enabled us to differentiate between nitrogen uptake and utilization efficiency. For both parameters we observed significant differences between hybrid and inbred varieties. Analysis of inter-trait relationships among more than 20 morpho-physiological traits enabled us to elucidate the key factors contributing to superior NUE and a positive agro-ecological balance in modern hybrids.

W273: Engineering NUE

TOND1, a Novel Gene Improves Tolerance to Nitrogen Deficiency in Rice

Chuanqing Sun, China Agricultural University, Beijing, China

Nitrogen (N), the most important mineral nutrient for plants, is critical to agricultural production systems. An N deficiency could severely affect rice growth and decrease rice yields. However, the excessive use of N fertilizer has caused severe pollution to the agricultural and ecological environments. Breeding crops that require less input of N fertilizer has been called for in the 'Second Green Revolution'. Here, we identified a major quantitative trait locus on chromosome 12, *Tolerance Of Nitrogen Deficiency 1* (*TOND1*), which confers tolerance to N deficiencies in rice of the *indica* cultivar Teqing. Sequence verification of 150 rice cultivars (75 *indica* and 75 *japonica* cultivars) from 18 countries and regions demonstrates that only 27.3% cultivars (41 *indica* cultivars) contain *TOND1* whereas 72.7% cultivars, including the remaining 34 *indica* cultivars and all 75 *japonica* cultivars, did not harbor the *TOND1* allele. Over-expression of *TOND1* has increased the tolerance to N deficiency

in the *TOND1-deficient* rice cultivars. The identification of *TOND1* provides a molecular basis for breeding rice varieties that improve grain yield with decreased N fertilizer input.

W274: Engineering NUE

Transgenic Poplar Expressing Pine GS1a Show Alterations in Nitrogen Homeostasis During Drought

Edward Kirby, Rutgers University, Newark, NJ

Transgenic hybrid poplars engineered to express ectopically the heterologous pine cytosolic *GS1a* display significant pleiotropic phenotypes including enhanced growth, enhanced nitrogen use efficiency, and resistance to drought stress. The present study was undertaken in order to assess mechanisms whereby ectopic expression of pine *GS1a* in transgenic poplars results in enhanced agronomic phenotypes. Microarray analysis using the Agilent *Populus* whole genome array allowed identification of genes differentially expressed between wild type (WT) and GS transgenics in four tissues (sink leaves, source leaves, stems, and roots) under three growth conditions (well-watered, drought, and recovery). Analysis revealed that differentially expressed genes in functional categories related to nitrogen metabolism show a trend of significant down-regulation in GS poplars compared to the WT, including genes encoding nitrate and nitrite reductases. Down-regulation of these genes was verified using qPCR, and downstream effects were tested using NR activity assays. Results suggest that higher glutamine levels in GS transgenics regulate nitrate uptake and reduction. Transcript levels of nitrogen-related genes in leaves, including GS/GOGAT cycle enzymes, aspartate aminotransferase, GABA shunt enzymes, photorespiration enzymes, asparagine synthetase, phenylalanine ammonia lyase, isocitrate dehydrogenase, and PII, were also assessed using qPCR revealing significant differences between GS poplars and the WT. Moreover, metabolites related to these differentially expressed genes showed alterations, including higher levels of GABA, hydroxyproline, and putrescine in the GS transgenic. These alterations in nitrogen homeostasis offer insights into mechanisms accounting for drought tolerance observed in GS poplars.

W275: Engineering NUE

NLP Transcription Factors Governing Nitrate-Responsive Gene Expression

Shuichi Yanagisawa, The University of Tokyo, Tokyo, Japan

Nitrate is the most abundant inorganic form of nitrogen in soils and a major source of nitrogen for plants. As a signaling molecule in plants, nitrate also plays critical roles in regulating a variety of physiological processes, including expression of nitrogen assimilation-related genes. We previously identified a nitrate-responsive *cis*-element (NRE) that is conserved among promoters of nitrite reductase gene from many plant species (1,2). Recently, we found that NIN-like proteins (NLPs) bind to the NRE and function as transcriptional activators that mediate nitrate signal (3,4). NLPs directly regulated expression of genes encoding nitrogen assimilation-related proteins, including nitrate transporter, nitrate reductase, chroloplastic nitrite transporter and nitrite reductase, in Arabidopsis (3,5-9). Furthermore, by transcriptome analysis with transgenic Arabidopsis plants in which NLP activity was reduced, NLPs were found to control expression of most of nitrate-responsive genes directly or indirectly. Hence, NLPs were identified as master regulators for nitrate response (8). These may implicate that enhancing NLP activity leads to improvement of nitrogen utilization efficiencies (NUE) of plants. However, since NLP activity is post-translationally regulated by nitrate signal (4), contrivances are necessary for development of a new strategy for improvement of NUE with NLPs. I will discuss our future challenge to improve NUE with NLPs. (1) *Plant J.* 63:269, 2010. (2) *BBRC*, 411:708, 2011. (3) *Nat. Commun.* 4:1617, 2013. (4) *Plant Signal Behav.* 8: e25975, 2013. (5) *PCP* 52:824, 2011. (6) *SSPN* 59:612, 2013. (7) *PCP* 55:1311, 2014. (8) *JXB*, 65:5589, 2014. (9) *Plant Sci.* 229:167, 2014.

W276: Engineering NUE

Towards Synthetic Nitrogen-Fixing Symbioses in Grasses

Michael K. Udvardi[1], Evangelia Kouri[1], Jean-Michel Ané[2], Kevin Garcia[3], John Peters[4], Amaya Garcia Costas[4], Florence Mus[4], Chris Voigt[5], Min-Hyung Ryu[5], Giles Oldroyd[6], Ponraj Paramasivan[6], Philip Poole[7], Barney Geddes[7] and Ramakrishnan Karunakaran[6], (1)The Samuel Roberts Noble Foundation, Ardmore, OK, (2)Department of Agronomy - University of Wisconsin Madison, Madison, WI, (3)University of Wisconsin-Madison, Madison, WI, (4)Montana State University, Bozeman, MT, (5)Massachusetts Institute of Technology, Cambridge, MA, (6)John Innes Centre, Norwich, United Kingdom, (7)University of Oxford, Oxford, United Kingdom

Nitrogen fertilizers fuelled the Green Revolution and today inject over 100 million tons of reactive-N per year into agricultural systems. Without N-fertilizers there would be 2 billion fewer people alive today, yet massive use of fertilizers over much of the globe is compromising human health and natural ecosystems, and challenging the sustainability of modern agriculture. In contrast, millions of resource-poor farmers lack sufficient N-fertilizer to ensure good harvests, especially in Africa where yields are often only 10-20% of yield potential for staples like maize. As a step towards solving these contrasting N-related problems, we aim to build synthetic nitrogen-fixing symbioses between bacteria and grasses, based on knowledge gained from decades of research on natural nitrogen-fixing symbioses in legumes. Key steps in this synthetic biology project include engineering of: signal compound production in bacteria and signal recognition in plants; concomitant biosynthesis of a specialized C-source by the plant for use by the bacteria; catabolism of this specialized C-source for energy production, as well as nitrogen fixation, respiratory protection of nitrogenase, and conditional suppression of ammonia assimilation in bacteria; and, finally, ammonium uptake by plant cells. Chassis' for the bacterial synthetic biology are natural endophytes or epiphytes of grasses, while the target model and crop species are barley and maize. Significant progress has been made for several of these steps and will be presented at the NUE workshop. Ultimately, substantial synthetic associative nitrogen-fixation in staple food crops could increase yields of resource-poor farmers and decrease the need for industrial N-fertilizers in resource-rich agricultural systems.

W277: EPIC: the Plant Epigenome Project

Epigenetic Inheritance and the Epigenome in Plants

Rob Martienssen, HHMI-GBMF Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

W278: EPIC: the Plant Epigenome Project

A One Precursor One siRNA Model for Pol IV-Dependent siRNA Biogenesis
Steve Jacobsen, University of California at Los Angeles, Los Angeles, CA
RNA-directed DNA methylation in Arabidopsis thaliana is driven by the plant-specific RNA Polymerase IV (Pol IV). It has been assumed that a Pol IV transcript can give rise to multiple 24-nt small interfering RNAs (siRNAs) that target DNA methylation. Here, we demonstrate that Pol IV-dependent RNAs (P4RNAs) from wild-type Arabidopsis are surprisingly short in length (30 to 40 nt) and mirror 24-nt siRNAs in distribution, abundance, strand bias, and 5' -adenine preference. P4RNAs exhibit transcription start sites similar to Pol II products and are featured with 5' -monophosphates and 3' -misincorporated nucleotides. The 3' -misincorporation preferentially occurs at methylated cytosines on the template DNA strand, suggesting a co-transcriptional feedback to siRNA biogenesis by DNA methylation to reinforce silencing locally. These results highlight an unusual mechanism of Pol IV transcription and suggest a ‘‘one precursor, one siRNA’’ model for the biogenesis of 24-nt siRNAs in Arabidopsis.

W279: EPIC: the Plant Epigenome Project
The Argonaute-binding platform of NRPE1 evolves through modulation of intrinsically disordered repeats
Joshua T Trujillo, University of Arizona, Tucson, AZ, Mark A Beilstein, School of Plant Sciences, University of Arizona, Tucson, AZ and **Rebecca A. Mosher**, The University of Arizona, Tucson, AZ
Argonaute proteins are important effectors in RNA silencing, but they must interact with other cellular machinery to cause silencing. Ago hooks are a conserved linear motif responsible for interaction with Argonaute proteins, but little is know about the sequence surrounding these small motifs, which must restrict or enable interaction with specific Argonautes. Here we investigated the evolutionary dynamics of an Ago-binding platform in NRPE1, the largest subunit of RNA Polymerase V. We compared NRPE1 sequences from more than 50 species, including dense sampling of two plant lineages. Our study demonstrates conservation of intrinsic disorder and repetitive character within the Ago-binding platform, but loss of sequence conservation. We reveal that loss of conservation is due to frequent expansions and contractions within tandem repeats, along with relaxed selection. Consequences and possible evolutionary drivers of this diversity will be discussed.

W280: EPIC: the Plant Epigenome Project
Investigating Chromatin Structure at Multiple Scales in Maize
Hank W. Bass[1], Daniel L. Vera[1], Eli Rodgers-Melnick[2] and Edward S. Buckler[3], (1)Florida State University, Tallahassee, FL, (2)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (3)USDA-ARS-Cornell University, Ithaca, NY
We have investigated genomic and epigenomic features of maize chromatin at several levels using computational, biochemical, and cytological approaches. Recent findings from several studies will be presented. These include [1] computational identification and characterization of motifs capable of forming non-duplex, G-Quadruplex (G4) DNA; [2] chromatin structure profiling using Differential Nuclease Sensitivity (DNS-seq) mapping; and [3] 3D molecular cytology of spatiotemporal patterns of DNA replication. At the DNA level, maize chromosomes were shown to harbor thousands of G4 DNA motifs with hot-spots located on the template strands in the 5’ UTR and the beginning of the first intron. Many of these genes function in metabolic signaling pathways responsive to low energy states such as hypoxia. At the chromatin level, nuclease sensitivity profiles highlight functional genomic regions. MNase hypersensitive (HS) sites at promoters predict transcriptional activity while intergenic MNase HS sites explain a large proportion of heritable phenotypic variation in structured populations. DNS-seq maps biochemically defined epigenomic features associated with multiple types of open chromatin. At the level of higher-order chromatin structure and nuclear architecture, we examined spatiotemporal patterns of DNA replication at early, middle, and late S-phase of EdU-pulse labeled nuclei from developing maize root tips. Maize euchromatin was found to exist as an intermingled but distinct mixture of two components/compartments, distinguished by their condensation state and replication timing (early versus middle S) in mitotic and endocycling nuclei. Together, these findings will help bridge the genotype-phenotype gap with integrative information and annotations that span multiple scales of epigenomic resolution.

W281: EPIC: the Plant Epigenome Project
The EPIC CoGe Browser
Eric Lyons, University of Arizona, Tucson, AZ, Mark A Beilstein, School of Plant Sciences, University of Arizona, Tucson, AZ and Brian D. Gregory, University of Pennsylvania, Philadelphia, PA
EPIC-CoGe is the extension to CoGe that provides several key features of data management and analysis to CoGe that was funded by the Gordon and Betty Moore Foundation. These features include permitting individual researchers to integrate and manage new genomes, keep them private, and share them with collaborators; integrate funcitonal (RNASeq, epigenomic) and diversity (SNPs, variants) data, and visualize them in JBrowse. This work has lead to the first read/write genomics platform and in the past two years EPIC-CoGe has been available, researchers have added over 5000 new genomes and 5000 new functional and diversity datasets. As part of a NSF funded project classifying and analyzing the evolution of plant long non-coding RNAs, EPIC-CoGe will continue to expand. This talk will provide a summary of EPIC-CoGe’s features and planned future development. EPIC-CoGe is available at: http://genomevolution.org.

W282: Equine 1
Implications of ENCODE Regulatory Maps for the Genetics of Complex Traits
Matthew T. Maurano, Institute for Systems Genetics, NYU School of Medicine, New York, NY

W283: Equine 1
UCSC Genome Browser: Platform for Data Display and Integrated Analysis
Robert Kuhn, UC Santa Cruz, Santa Cruz, CA
The UCSC Genome Browser provides a uniform display platform for annotations on any organism, hosting more than 150 assemblies of 90 animals as native data sets and providing a framework for any user to add an organism of choice. User data in a variety of formats may be displayed alongside UCSC-resident data and intersected with it.

Existing resources on the horse assembly, for example, include Horse RefSeq Genes, mappings of human proteins to the horse genome, and alignments of all genbank mRNAs that pass the homology threshhold from any species to horse. These may be intersected with user-generated quantitative data, such as RNA-seq, signal strength for histone modification ChIP-seq data or other epigenetic features. User-generated datasets may also be correlated to each other. Comparative genomics resources allows leverage of information from other animals and make it easy to navigate to orthologous regions on those other animals' browsers.

W284: Equine 1

Podium: Functional Organization and Inheritance of Satellite-Less Equid Centromeric Domains

Elena Giulotto[1], Solomon Nergadze[1], Federico Cerutti[1], Riccardo Gamba[1], Francesca M. Piras[1], Marco Corbo[1], Claudia Badiale[1], Eleonora Cappelletti[1], Josef McCarter[2], Douglas Antczak[3], Donald Miller[3], Kevin Sullivan[2] and Elena Raimondi[1], (1)University of Pavia, Pavia, Italy, (2)National University of Ireland, Galway, Galway, Ireland, (3)Cornell University, Ithaca, NY

The centromere is the site of kinetochore assembly required for correct chromosome segregation during cell division. At centromeres, despite the evolutionary conservation of proteins, DNA sequences are highly variable, even among related species. This paradox is now explained by the knowledge that the centromeric function is epigenetically specified.

This chromatin domain has so far escaped comprehensive molecular analysis due to its typical association with highly repetitive DNA (satellite DNA). Although satellite DNA is a common feature of mammalian centromeres, we proved that several equid centromeres are completely satellite-free, thus representing a unique model for studying this epigenetically specified locus.

In previous work, the functional organization of the single-copy centromere of horse chromosome 11 was analysed by ChIP-on-chip with an antibody against CENP-A, the centromeric histone-H3 variant. Inter-individual positional variation of CENP-A domains was observed, giving rise to “epialleles" and proving that centromeric domains are autonomous relative to the underlying DNA and are characterized by positional instability.

ChIP-seq experiments with anti-CENP-A antibodies were performed in Grevy's zebra and donkey. The sequences of several satellite-less centromeres from these species were *de novo* assembled, allowing us to perform a comparative analysis between centromeric and orthologous non-centromeric DNA. The inheritance of centromeric domains in hybrid families (horses, donkeys and mules) was also analysed. The possible role of DNA breakage, methylation, transcription and heterochromatic marks in the establishment and function of the centromere will be discussed

The satellite-less equid centromeres represent a powerful system to study the molecular organization and evolution of mammalian centromeres.

W285: Equine 1

Podium:Identifying Genomic Regions Undergoing Selection in Racehorses

Felipe Avila, University of Minnesota, St Paul, MN

Intense breeding for speed over a relatively small number of generations has resulted in elite athletic performance within Thoroughbred, Quarter Horse and Standardbred racehorses. The goals of this study were to: 1) identify the genomic regions undergoing selection in these three breeds; 2) investigate haplotype sharing among racehorse breeds; and 3) identify putative candidate genes that contribute to the racing phenotype.

Five hundred and forty-six Standardbreds, 519 Thoroughbreds, and 24 Quarter Horses were genotyped at, or imputed up to, 2 million SNP markers distributed across the genome. Genomic regions harboring signatures of selection were identified using two F_{ST}-based statistics (*di* and hapFLK), and local haplotype sharing among individuals of each breed was calculated using the hapQTL software. Genomic regions of interest (ROIs) were defined as those in which significant *di* windows (99^{th} percentile of the empirical distribution) and hapFLK values (-log10[p-value] > 3) overlapped with significant SNP markers found by hapQTL (-log10 [Bayes factor] > 5). Seventy-five ROIs, distributed across 22 autosomes and the X chromosome, averaging between 10kb and 1.5Mb in length, were prioritized for further investigation. Ninety-two equine transcripts were identified within ROIs, retrieved using the BioMart software, and investigated for their association with racing ability. Pathway analysis was performed to identify protein-coding genes significantly associated with main metabolic pathways that might contribute to the racing phenotype.

Next, WGS data from these three breeds will be used for the discovery and annotation of biologically relevant allelic variants that might account for their increased racing ability.

W286: Equine 1

Podium: The Role of Myostatin on the Conformation and Gaits of the Icelandic Horse

Liesbeth Francois, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

The influence of myostatin (MSTN) on conformation, an important selection criterion in many horse breeds, has only recently come to light. Although conformation is considered a complex trait influenced by multiple genes, recent studies have shown that MSTN not only influences the morphological type of an individual, but also the individual’s performance. Through its role as a repressor in the development and regulation of skeletal muscle mass, MSTN is a major gene of interest. This is a first study looking into its influence on conformation and riding ability of the Icelandic horse, a breed known for its robust and compact conformation, and additional gaits, tölt and pace. Three SNPs (g.65868604G>T, g.66493737C>T and g.66495826A>G) within MSTN were analyzed in 195 Icelandic horses and their association to official estimated breeding values for 16 traits was evaluated. Significant associations ($p<0.05$) were found with several conformation traits such as neck/withers/shoulders, hooves, leg stance and total conformation. Furthermore, there was an indication that myostatin possibly influences the performance of the ambling trait tölt, a gait unique to gaited horse breeds ($p=0.07$). This study showed that MSTN plays a role in the complex background of breed-specific conformation traits and has a possible influence on tölt, a unique and highly selected trait within this breed. Further analysis will help to decipher the specific pathway by which MSTN influences these traits.

W287: Equine 1

Podium: Improving Annotated Gene Structure in the Horse with a Publically Available Transcriptome of Six Tissues

Tamer A. Mansour, School of Veterinary Medicine, UC Davis, Davis, CA

The current annotation of the horse genome is inadequate for successful genome to phenome analyses. There is a paucity of evidence codes, and many inconsistencies exist between publicly available annotations from the NCBI and Ensembl databases. Additionally, the lack of tissue-specific transcription profiles creates an obstacle for modern functional genomics and system biology approaches. The horse genome also contains confounding frameshift errors that add another barrier for proper gene prediction. In this study, we used Tophat and Cufflinks to produce reference-guided assemblies for 69 samples with ~1.5 billion Illumina sequencing fragments. We merged the assemblies into six tissue-specific transcription profiles for cerebellum, brainstem, spinal cord, retina, muscle and skin. The final merger of all assemblies overlaps with 63% and 73% of NCBI and Ensembl loci, respectively, capturing about 72% and 81% of their coding bases. Comparing our assembly to the most recent transcriptome annotation shows ~85% overlapping loci. In addition, at least 40% of our annotated loci represent novel transcripts. As part of this work, we present a new application of digital normalization to allow for integrative analyses of RNAseq, which captures very low abundance transcripts. This enabled us to extend UTRs of our gene models and identify normally low-level transcribed ncRNA. Finally, we built a lightweight variant analysis pipeline using the GATK toolkit to detect and fix genome-sequencing errors affecting open reading frames. All of our transcriptome assemblies, a liftover of NCBI annotation and the corrected version of the horse genome are readily available as UCSC track hubs.

W288: Equine 1

Podium: RNA-SEQ Analysis of the Exercise in Horses: Insights on Transcribed Exons, Introns and Repeats

Stefano Capomaccio, Sport Horse Research Centre - University of Perugia, Perugia, Italy

The horse is probably the best animal model for investigating genomic response to exercise-induced stress, due to its natural aptitude for athletic performance and the relative homogeneity of its genetic background. We applied RNA-seq to PBMCs of six 3 years old race horses sex matched collected at rest and after a 2000-meter competition. We observed a transcription shift from coding to non-coding regions therefore we separately analyzed exon and intron compartments. In both analyzed compartments, network and GO analysis revealed mechanisms known to be activated by stress as well as functions to preserve energy devoted to other processes.

A large number of transcripts, corresponding to intergenic and intronic regions associated with new transcriptional elements, were identified. This data might be correlated with transcriptional activity related to nascent transcription, co-transcriptional splicing events or transcription of long noncoding RNAs or enhancer RNAs that are know to contribute to changes in intronic read counts.We observed a post-race increase of reads mapping to repeats, especially to LINE1 in intergenic and intronic regions.

Our results reinforce the hypothesis that transposable elements (TE) and intronic sequences may serve as transcriptional units capable of enriching transcriptomes with limited genomic resources, such under stress conditions. We also found that 9 full-length LINE1 elements are up-regulated after the race. This suggests that a great effort induced by exercise may - in principle - activate LINE1 retrotransposition, as already demonstrated in human and mouse tissues and in certain sporadic cancers.

W289: Equine 1

Podium: Optimized Method for Extracting Circulating Small RNAs from Long-Term Stored Equine Samples

Alicja Pacholewska, Institute of Genetics, University of Bern, Bern, Switzerland

Small RNAs play an important role in shaping a cell's transcriptome profile. One miRNA may affect the expression of multiple target genes. Moreover, it has been suggested that circulating miRNA may act as hormones and be transported between cells. MiRNAs are promising biomarkers for many diseases, including cancer. MiRNA circulating in the blood are relatively easy to sample and their high stability enables implementation of simple procedures for future diagnostics. However, extracellular miRNA are only present in low concentrations and this still poses challenges regarding these procedures.

While small RNA extraction methods have been established for model organisms like human and mouse, little is known regarding optimized procedures for equine miRNA extraction. Species- specific differences in the concentration and composition of serum small RNAs have been reported.

In the presented study we optimized a method for the extraction of small RNA from equine serum. Moreover, six small RNA samples isolated with the presented protocol were then used for library preparation and successfully sequenced on an Illumina MiSeq. The sequencing data confirmed expression of some predicted equine miRNA sequences available in miRBase. The method also showed high efficiency with serum samples stored for many years.

Next, we optimized a total RNA (including small RNA) extraction method from frozen equine EDTA blood samples. In addition, we evaluated the stability of total RNA from frozen EDTA blood samples after long-term storage (> 10 years). We will present the procedure optimization steps and the most recent results at the PAG meeting.

W290: Equine 1

Podium: The Genetic Mechanisms Driving Cerebellar Abiotrophy in Arabian Horses

Erica Scott, School of Veterinary Medicine, UC Davis, Davis, CA

Equine cerebellar abiotrophy (CA) is a hereditary neurodegenerative disease affecting the Purkinje neurons of the cerebellum, resulting in an ataxic phenotype. It is inherited as an autosomal recessive trait and associated with a single nucleotide polymorphism (SNP) on equine chromosome 2 (CA SNP), which is located within a *TOE1* exon and in proximity to *MUTYH* on the opposite strand. Unraveling which gene and associated pathway the CA SNP is affecting may elucidate the molecular mechanism of CA. RNA-seq (100 bp PE strand-specific) was performed in cerebellar tissue of six CA-affected and five age-matched unaffected horses. Based on principle component analysis and dendogram plots, samples were clustered with four CA-affected and five CA-unaffected horses for further analyses. Two pipelines for analysis of differentially expressed genes were used: Tophat2/Cufflinks/Cuffdiff2 and Kallisto/edgeR. There were 195 significant differentially expressed genes in agreement between both analyses (FDR=0.05). *TOE1* (log(fold change)=0.51, q=0.618) and *MUTYH* (log(fold change)=0.03, q=0.992) were not among the differentially expressed genes. However, genes such as *CALB1* (log(fold-change)=-4.6, q<0.001) and *PCP2* (log(fold-change)=-3.26, q<0.001) that are specifically expressed in Purkinje neurons, were significantly under-expressed, as expected. The cerebellar transcriptome includes expression of over 18,000 genes, of which several have extended UTRs and overlapping transcriptional regions. This transcriptome annotation was essential in identifying other potential molecular mechanisms that may be affected by the CA SNP.

W291: Equine 1

Podium: Skeletal Variation in Tennessee Walking Horses Maps to the LCORL/NCAPG Gene Region

Elizabeth A. Staiger, Cornell University, Ithaca, NY

Conformation has long been a driving force in horse selection and breed creation as a predictor for performance. The Tennessee Walking Horse (TWH) ranges in size from 1.5 to 1.7 meters and is often used as a trail, show, and pleasure horse. To investigate the contribution of genetics to body conformation in the TWH, we collected DNA samples, body measurements, and gait/training information from 282 individuals. We analyzed the 32 body measures with a Principal Component Analysis (PCA). Principal Component (PC)1 captured 28.5% of the trait variance, while PC2 comprised just 9.5% and PC3 6.4% of trait variance. All 32 measures correlated positively with PC1, indicating that PC1 describes overall body size. 109 horses were genotyped using the EquineSNP70 bead chip and marker association assessed using PC1 scores as a phenotype. Mixed-model linear analysis (EMMAX) revealed a well-documented candidate locus on ECA3 (raw $p=3.86 \times 10^{-9}$) near the *LCORL* gene. A custom genotyping panel enabled fine-mapping of the PC1 body-size trait to the 3' end of the *LCORL* gene ($p=7.09 \times 10^{-10}$). This position differs from other reports suggesting SNPs upstream of the *LCORL* coding sequence regulate expression of the gene and therefore, body size in horses. Fluorescent *In Situ* Hybridization (*FISH*) analysis defined the position of a highly homologous 5 kb retrogene copy of *LCORL* (previously ChrUn of the EquCab 2.0 assembly) at ECA9 q12-q13. This is the first study to identify putative causative SNPs within the *LCORL* transcript itself, which are associated with skeletal size variation in horses.

W292: Equine 2

An Improved Horse Reference Genome for Enhanced Biological Discovery

Richard Green, University of California, Santa Cruz, Santa Cruz, CA

Horses are one of the most intensely studied domesticated animals. In the 5,500 years since domestication, humans have performed selective breading to produce a variety of horse breads with dramatically different phenotypic traits. Thus, horses are a useful model system for understanding the selective forces underlying domestication and the genetic architecture of specific morphological and behavioral traits. However, the horse genome - a necessary reagent for gene-mapping studies - is not as well assembled as most genetic model organisms. We used a new method for high-contiguity long-range genome assembly to produce a new, de novo reference horse genome assembly. We will discuss the technology behind this assembly and demonstrate the accuracy and increased utility of this genome for revealing the recent population history of horses.

W293: Equine 2

Update on Progress Toward EquCab3

Theodore S. Kalbfleisch, University of Louisville, Louisville, KY

In our last update, we reported the results of our latest assembly effort that augmented the original Sanger sequence data from EquCab2 (Wade et al.) with ~45X of Illumina PE reads from PCR free library preps, producing a contig N50 of 765kb. This is a nearly 7-fold increase in median contiguity over the current assembly. Since then, we have generated ~16X coverage using PacBio SMRT cell sequencing, as well as ~40X coverage with a Dovetail Genomics Chicago™ sequencing library. Here, we detail the latest N50, and annotation metrics for assembly as we incorporate this new sequence data, and describe our plans for moving forward with the reference over the next few months.

W294: Equine 2

Podium: Expansion of the Horsegene Database Enables Stronger Collaborations between Veterinarians and Geneticists

Brandon D. Velie, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

Funded by the EU's Seventh Framework Programme, the Horsegene project united the efforts of multiple European research groups that had been working independently on the genetics of disease in horses for many years. One initiative of the consortium was to better facilitate the sharing of information between researchers through the creation of the Horsegene Database (https://horsegene.interbull.org/). The database securely houses unlimited pedigree, phenotype, and genotype data and contains a web-interface that allows consortium members to clearly see what types of genetic (e.g. WGS, RNAseq) and phenotypic information are available. Interest from the global equine genomics community in expanding the use of the database beyond the Horsegene consortium has been strong. Further, equine genomics research is often limited by the ability of geneticists to effectively communicate with practicing veterinarians to attain adequate sample numbers for robust analyses. A new addition to the web-interface is targeted at increasing awareness of current equine genetic research efforts. The contact information of equine geneticists around the world, as well as the details of specific diseases and traits being investigated can now be found under the "Get Involved" option of the web-interface. The tab provides a breakdown of genetic research currently underway in horses and can be publically accessed. In conclusion, we have developed a valuable tool that will help to better facilitate research studies that use genetics and genomics to advance our understanding of the genetic mechanisms underlying phenotypic variation in horses.

W295: Equine 2

NRSP-8 Bioinformatics Update

James M. Reecy, Iowa State University, Ames, IA

W296: Evolution of Genome Size

Genome size evolution among closely related species in *Gossypium*

Corrinne E. Grover[1], William S. Sanders[2], Dinum Perera[3], Mark A. Arick II[4], Daniel G. Peterson[5], Jonathan F. Wendel[1], Jodi A. Scheffler[6] and Brian Scheffler[7], (1)Iowa State University, Ames, IA, (2)Mississippi State University, MS State, MS, (3)Mississippi State University, Starkville, MS, (4)Institute for Genomics, Biocomputing, and Biotechnology - Mississippi State University, MS

State, MS, (5)Mississippi State University, Mississippi State, MS, (6)USDA-ARS, Stoneville, MS, (7)USDA-ARS Genomics and Bioinformatics Research Unit, Stoneville, MS
Genome size is an extremely labile characteristic of plant genomes, often experiencing rapid changes over short evolutionary time. The size differences among and lability of plant genomes can largely be attributed to the dynamics of the repetitive sequences that populate and influence these genomes. Thus far, research into genome size evolution in *Gossypium* has largely focused on intergenomic comparisons of repetitive populations among more distantly related genome groups; however, recent analyses have suggested that while genome size within groups of closely related species (e.g., within the diploid cotton A-genome group) can be nearly identical (1667 versus 1698 Mbp in *G. herbaceum* and *G. arboreum*, respectively), the repetitive profile for each genome can vary substantially. Here we extend intragenomic repetitive profile analysis to the smallest clade of cotton species (D-genome), whose 13 known species range in genome size from 841 Mbp in *G. gossypioides* and *G. thurberi* to 934 Mbp in *G. laxum* and *G. lobatum*. Using a cluster-based analysis, we profile the repetitive content for each species and evaluate the evolution these in a phylogenetic context, providing insight into the evolution of repeats over short evolutionary distances and in the presence of introgression.

W297: Evolution of Genome Size

The Evolution of Polyploid Genomes in *Tragopogon* (Asteraceae): Changes in Gene Content, Gene Expression, and Karyotypes

Pamela S. Soltis and Douglas E. Soltis, University of Florida, Gainesville, FL
Tragopogon is a textbook model for studying allopolyploidy, with two recently (~80 years old; 40 generations in these biennials) and repeatedly formed natural allotetraploids, *T. mirus* and *T. miscellus*. These species originated during the first half of the 20th century from diploid parents (*T. dubius* and *T. porrifolius* and *T. dubius* and *T. pratensis*, respectively) that were introduced into North America from Europe. These two species have each originated multiple times during the past 80 years in the narrow geographic region of eastern Washington and northern Idaho, USA. Synthetic lines (S_0-S_5) of both allotetraploids have also been developed. In addition, the allotetraploid *T. castellanus* from Spain (0.5 - 1.5 million years old) provides an older polyploid for comparison. Moreover, ancient whole-genome duplication (WGD) also occurred deep in the history of Asteraceae. *Tragopogon* therefore presents research opportunities not available in other systems, providing the chance to investigate the patterns, tempo, mechanisms, and evolutionary forces driving post-WGD genome fractionation and diploidization. We will provide both new data and an overview of the changes in gene expression, homeolog loss, and chromosomal changes that have occurred repeatedly in the unique evolutionary system provided by young polyploids in *Tragopogon* and extend these patterns deeper in time via analyses of *T. castellanus* and the early history of the Asteraceae.

W298: Evolution of Genome Size

The Unique Composition of Transposons in the Basal Dicot Sacred Lotus

Ning Jiang, Michigan State University, East Lansing, MI
Transposable elements (TEs) are pervasive among eukaryotes and are often the largest component in these genomes. Their composition reflects the interaction between TEs and their host genomes, as well as the consequence of vertical inheritance and horizontal transfer. Sacred lotus, an ancient eudicot, represents the sister lineage to all core eudicots, and provides a unique reference for studying the evolution of dicots and the split between monocots and dicots. The sequencing of the sacred lotus genome has allowed us to characterize its repetitive content. Here, we report that 63% of the genome is composed of repetitive sequences, and the majority of them (55% of the genome) are identifiable TEs. Analysis of the TE content and diversity in sacred lotus revealed unique composition of TEs compared to other plant genomes characterized so far. Among LTR elements, a considerable portion (15%) are associated with non-canonical LTR ends which has not been reported for any other organisms. The detection of thousands of LTR elements with non-canonical ends suggests the long-term co-evolution between integrase and the termini of the element, providing new guideline for future annotation of LTR elements. Sacred lotus also contains the highest coverage and copy number of *hAT* elements among all genomes sequenced to date. The *hAT* elements make up a substantial portion (9%) of the genome, which may have favored the generation of novel domesticated *hAT* genes. In addition, the genome contains 1447 Pack-MULEs and provides the first evidence for the acquisition preference of GC-rich sequences by Pack-MULEs outside monocots.

W299: Evolution of Genome Size

A Mechanism for Genome Size Reduction Following Chromosomal Rearrangements

Longhui Ren, Iowa State University, Ames, IA and Steven B. Cannon, USDA-ARS-CICGRU, Ames, IA
A genomic inversion occurs when a segment of chromosome detaches by breakage and reinserts or reattaches in a reversed orientation. Several mechanisms can introduce inversions in a genome. Studies show that inversions can result in speciation and X chromosome evolution, probably through suppression of recombination between incompatible karyotypes. Previous research on *Arabidopsis thaliana* and *Arabidopsis lyrata* indicates that genome rearrangements are associated with genome shrinkage in *A. thaliana*, where rearranged regions are on average shorter in *A. thaliana* than are collinear but unrearranged (ancestral) regions in *A. lyrata*. We compare the genomes of two wild ancestors of peanut, *Arachis duranensis* and *Arachis ipaensis*, whose genomes merged several thousand years ago in a rare genetic event. These two ancestral diploids separated from each other about 3 million years ago, which makes these interesting models of genomic evolution. Several large inversions occurred in the *A. duranensis* genome, and lead to the contraction of genome through net removal of transposons and locally duplicated genes in inverted regions. This is evident when comparing size ratios of syntenic blocks in inverted and un-inverted regions, which have smaller syntenic block size in those inverted regions in *A. duranensis*. Our research proposes a possible mechanism of genome size changes, which needs to be tested with other species to confirm our model and to uncover the details in this mechanism.

W300: Evolution of Genome Size

Evolutionary Patterns of Chromosomes and Genomes

Jianming Yu[1,2], Xianran Li[2] and Michael Scanlon[3], (1)Iowa State University, Ames, IA, (2)Department of Agronomy, Iowa State University, Ames, IA, (3)Cornell University, Ithaca, NY

Our understanding of genome and chromosome evolution can be significantly improved if patterns of such evolution can be discovered across taxonomic groups and species with varied complexity. In the first study, we demonstrated that variation in chromosome size for 886 chromosomes in 68 eukaryotic genomes can be viably captured by a single model. This conserved boundary of chromosome-size variation indicates that cellular, molecular, and evolutionary mechanisms confine the chromosome lengths around a species-specific average chromosome length. In the second study, we uncovered a clear separation pattern of base-composition values calculated across polymorphic sites between basal and derived populations separated by a bottleneck event. DNA repair genes were found to be significantly enriched within genomic regions underlying the divergence of this genome phenotype. Our findings highlighted the need for global analysis of genomic data and the need for integration of molecular mechisms and evolution.

W301: Exploring Phytobiomes

The Vital Importance of Soil Health: from Laboratory Systems to Ecosystems

Leland J. Cseke, The University of Alabama in Huntsville, Huntsville, AL

Healthy soil is an important resource in both natural and agricultural ecosystems. It harbors both chemical and biological sources of nutrients that are required for plant growth, reproduction, insect and pathogen resistance, and response to changing environments. However, current agricultural practices rely heavily on the meticulous management of chemical nutrients. In natural ecosystems, nutrients are taken up by microorganisms and delivered to plants during interactions with plant roots in exchange for fixed carbon. While such symbiotic relationships play an essential role in the regulation of soil nutrient cycling and subsequent carbon management, the complexity and variation of such natural systems has hindered the accurate assessment of the factors that signal, establish and maintain these interactions. Thus, our lab has made use of simplified *Populus-Laccaria-Pseudomonas* laboratory systems and multiple "omics" approaches to integrate RNA-Seq, SWATH proteomics, ChIP-Seq, biochemical and FTIR imaging analyses to enhance our understanding of how atmospheric carbon is sequestered as plant and/or subsurface fungal biomass during symbiotic interaction and/or nutrient limitation. Our results indicate that molecular events at the protein level are key in drawing conclusions for such processes. We aim to make use of these laboratory systems to study natural ecosystems by providing a mechanism to probe how nature so effectively controls availability of soil nutrients through the activity of beneficial microbes. We discuss how holistic system design approaches, combined with modern molecular techniques, provide technologies for the deep repair of ecosystems and the development of solutions for depleted topsoil, broken hydrological cycles, and falling soil fertility.

W302: Exploring Phytobiomes

Harnessing the Microbiome for Agricultural Sustainability in Bioenergy-based Systems

Kelly Craven, Noble Foundation, Ardmore, OK

Declining reserves of mineral phosphorus and growing economic and environmental costs associated with fertilizer use (and misuse) have necessitated efforts to identify cropping systems and strategies that can be sustained under a low-input strategy. One approach to ameliorate such losses is to utilize microbial symbionts that have evolved to promote plant growth through nutrient and water acquisition as well as reduce plant stress when grown on marginal, low-quality soils. Soils such as these are expected to be tapped to grow cellulosic feedstocks for biofuel production. Here, we describe our efforts to maximize the performance and abiotic stress tolerance of switchgrass, a C4 grass native to the prairies of northern OK, through microbial symbiosis. Strain discovery combined with the implementation of high-throughput screens for potentially useful traits, have resulted in a manageable number of bacterial and fungal endophytes that we are testing in greenhouse trials. Results to date suggest that both biomass and drought tolerance can be enhanced by a novel type of mycorrhizae and bacteria have been identified that are being tested for phosphorus solubilization, nitrogen fixation and the alleviation of ethylene-induced plant stress.

W303: Exploring Phytobiomes

Analyzing the Leaf Microbiome across 270 Diverse Maize Lines

Jason G Wallace[1], Karl A. Kremling[2], Shu-Yun Chen[2], Mei-Hsiu Su[2], Jeremy D. Pardo[2], Nicholas K. Lepak[3], Joshua S. Budka[3] and Edward S. Buckler[4], (1)University of Georgia, Athens, GA, (2)Cornell University, Ithaca, NY, (3)USDA-ARS, Ithaca, NY, (4)USDA-ARS-Cornell University, Ithaca, NY

A largely unexplored area in global food security research is the interaction between crop plants and the microbial communities around them. The exceptions usually involve either diseases (e.g., rusts and blights) or a few specific symbioses (especially rhizobia and mycorrhiza). The general microbial community—the crop “microbiome”—nonetheless plays an important role in crop health. To better understand interactions between maize and its microbiome, we collected leaf samples from across ~270 diverse maize inbred lines to identify the genetic mechanisms that shape its microbial community. Community makeup was determined by targeted 16S amplification and deep sequencing. Several bacterial taxa show high heritabilities. Genome-wide association (GWAS) of these taxa identifies the major genetic loci influencing their abundance, and we identify recurring patterns across multiple associations. This is the most diverse maize microbiome study to date, and these results will direct future analyses to identify ways to manipulate the leaf microbiome to improve crop performance.

W304: Exploring Phytobiomes

The Influence of the Endophytic Grapevine Microbiome on Pierce’s Disease Development

Caroline Roper[1], James Borneman[1], Paul Ruegger[1], Jiue-In Yang[2] and Philippe Rolshausen[1], (1)University of California, Riverside, Riverside, CA, (2)National Taiwan University, Taipei, Taiwan

Pierce’s Disease (PD) of grapevine, caused by the xylem-limited bacterium *Xylella fastidiosa* (*Xf*), is a major threat to the grapevine industry. In vineyards that are under high PD pressure, there are interesting examples of vines exhibiting either no symptoms or very mild PD symptoms (disease-escaped). These differences are likely not attributed to the genetics of the plant because all vines in a vineyard are clonal. We hypothesize that the microorganisms inhabiting the xylem in these disease-escaped vines are inhibitory to *Xf* and subsequently reduce disease severity, due to their shared ecological niche. We have characterized the microbial communities residing in PD-infected vines and compared them to disease-escaped vines and identified a subset of beneficial organisms that are antagonistic to *Xf*. We characterized the fungal and bacterial endophytic communities using an Illumina MiSeq platform targeting the ITS and 16S rRNA genes, respectively. Pseudomonadales was

the most abundant bacterial taxonomic group, and Pleosporales was the most abundant fungal taxonomic group. Some bacterial and fungal phylotypes correlated either positively or negatively with PD severity. Furthermore, we cultured a subset of the endophytic microbes that possessed strong anti-*Xf* properties and suppressed PD symptom development in greenhouse bioassays. We are also currently assessing the role that members of the grape microbiome play in stimulating plant immunity. We envision harnessing these microbes to construct a beneficial synthetic phytobiome that can be deployed into grapevines during the nursery propagation process.

W305: Exploring Phytobiomes

High Throughput Phenotyping for Complex Traits : Case Study for Nitrogen Response in Wheat Based on the PhénoBlé Project

David Gouache[1], Benoit de Solan[2], Antoine Fournier[3], Alexis Comar[4], Fred Baret[5], Fabien Cormier[6], Stéphane Lafarge[7], Agathe Mini[8], Benoit Piquemal[1], Sébastien Praud[6], Jacques Le Gouis[9] and Katia Beauchene[3], (1)ARVALIS - Institut du végétal, Boigneville, France, (2)ARVALIS - Institut du végétal, Avignon, France, (3)ARVALIS - Institut du végétal, Ouzouer le marché, France, (4)HiPhen, Avignon, France, (5)INRA, Avignon, France, (6)BIOGEMMA, Chappes, France, (7)Biogemma, Chappes, France, (8)ARVALIS - Institut du végétal, Chappes, France, (9)INRA GDEC, clermont ferrand, France

Crop response to abiotic stress is a complex trait, muddled by genotype by environment interactions, and multiple underlying traits. This is typically the case for response to nitrogen in wheat. High throughput phenotyping provides access to intermediate level traits that can help in understanding, screening, and ultimately ameliorating nitrogen response. We provide results on the use of proximal remote sensing technologies used to investigate the response of an elite panel of French bread wheats to nitrogen, obtained via a project entitled PhénoBlé. These results have allowed us to identify certain remote sensing proxies for radiation interception and radiation use efficiency as promising traits for screening germplasm response to nitrogen. Based on these results, as well as multilocal trials screening the same panel for nitrogen response, we propose avenues for implementing these technologies, with a focus on plant breeding. We also present technological evolutions from the initial prototype system to facilitate wide adoption. Finally, we conclude by drawing parallels with the requirements for investigating complex crop-phytobiome interactions for improved tolerance to stresses.

W306: Exploring Phytobiomes

Roadmap for Phytobiomes Research and the International Phytobiomes Consortium

Jan E. Leach, Colorado State University, Fort Collins, CO, Kellye Eversole, Eversole Associates, Bethesda, MD and Gwyn A. Beattie, Iowa State University, Ames, IA

Phytobiomes are all of the living organisms in, on, and around plants, and encompass the many organisms that influence or are influenced by the plant or the plant environment, including the soil. Phytobiomes, thus, consist of other plants, animals (insects, nematodes and amoeba), and a wide diversity of microbes (viruses, bacteria, fungi, oomycetes, and algae), soil, and the environment. Modern technologies, such as high-throughput sequencing, computational biology, and many '-omics' technologies, are enabling exploration of the composition, function, and activities of phytobiomes. The application of these technologies has illustrated a vast potential for exploiting phytobiomes to increase crop production sustainably and improve agroecosystem health. The Phytobiomes Initiative promotes fundamental and applied research to gain a comprehensive, systems-level understanding of phytobiomes that can be used to improve crop production, quality, and safety. With input from a wide range of stakeholders, a draft roadmap for phytobiomes research and translation has been developed. The International Phytobiomes Consortium is a new, public-private collaborative organization focused on coordinating and advancing applied and fundamental phytobiomes research to ensure a sufficient supply of food, feed, and fiber. The roadmap and the new public-private consortium will be presented.

W307: Flax Genomics

BioNano Optical Map Improves the Flax Reference Genome

Sylvie Cloutier[1], Michael Deyholos[2], Raja Ragupathy[3], Tingting Zhu[4], Pingchuan Li[5], Ming-Cheng Luo[4] and Frank M. You[5], (1)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (2)University of British Columbia, Kelowna, BC, Canada, (3)Department of Plant Science, University of Manitoba, Winnipeg, MB, Canada, (4)Department of Plant Sciences, University of California, Davis, Davis, CA, (5)Agriculture and Agri-Food Canada, Morden, MB, Canada

The flax genome, first sequenced in 2012 using short reads, was assembled into 3,852 scaffolds larger than 1Kb covering a total of 300Mb. A number of genomic resources were developed to assign, order and orient the scaffolds by chromosomes. These included a physical map generated from 43K BAC clones assembled into 419 FPC contigs, 54Mb of BAC-end sequences, three genetic maps with >100K SNPs and 770 SSRs and the resequencing of 407 flax accessions which yielded 1.7M SNPs. Using these resources, a total of 768 scaffolds were refined and assigned to 15 linkage groups. To further improve the assembly, we recently constructed an optical map using BioNano technology. A total of 82Gb of raw DNA molecules were obtained representing more than 200x genome equivalent with N50 of 249 Kb. These raw molecules were assembled into 251 contigs with an improved N50 of 2.15 Mb and a total size of 317 Mb. The optical map was particularly useful for orienting scaffolds of the reference assembly. To date, a total of 286 Mb of scaffold sequences were anchored to 317 Mb BioNano map, covering 97% of all predicted genes of the WGS assembly. The draft pseudomolecules contained ~330 Mb of sequences (including gaps) with linkage groups varying in sizes from 13.8 to 25.1 Mb.

W308: Flax Genomics

Genome Mapping in the Flax Rust Fungus *Melampsora lini* Reveals Novel Avirulence Genes

Peter Dodds[1], Arwen Zhang[1], Claire Anderson[2], Nadya Farrah[2], Laura Rolston[2], Maud Bernoux[1], Simon Williams[3], Wenjie Wu[2], Adnane Nemri[1], Narayana Upadhyaya[1], Jeffrey G. Ellis[1], Bostjan Kobe[3], Adrienne Hardham[2] and David Jones[2], (1)CSIRO Agriculture, Canberra, Australia, (2)ANU, Canberra, Australia, (3)University of Queensland, Brisbane, Australia

Rust fungi such as *Melampsora lini* (flax rust) and *Puccinia graminis fsp tritici* (wheat stem rust) form specialised haustoria structures during infection that serve as nutrient uptake sites as well as delivering effector proteins to the host cell. We have been investigating the molecular basis

and cellular location of the recognition events between R and Avr proteins, as well as characterising Avr protein structure and function. Four avirulence (Avr) loci are known in flax rust and encode effectors that are recognised by host nucleotide-binding and leucine-rich repeat (NB-LRR) resistance proteins. Recently we generated a draft genome sequence of 190Mbp for flax rust represented in 21,000 contig sequences. Using RADseq, we have scored ~13,000 SNP markers in an *M. lini* F2 family to generate a large scale linkage map that has allowed anchoring of 67% of the genome sequence onto 28 linkage groups. This process identified physical regions co-segregating with several Avr genes and one avirulence inhibitor locus. Two new Avr genes have been cloned and confirmed by expression in resistance flax lines. Gene expression profiling during rust infection of flax reveals a common expression pattern for a subset of effector candidates that includes all of the known *Avr* genes.

W309: Flax Genomics

Genome Editing in Flax in Response to Environmental Stress

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 variations do not appear to be due to either random mutations or movement of transposable elements, but rather that the genome appears to be able to switch between two well-defined, different sequences at many loci. The genomes of eight lines, (genotrophs) all derived from the same progenitor line (Pl), have been compared, by whole genome sequencing, to the reference genome of the flax variety Bethune. The differences fall into two main classes. One, where regions of the genome have insertions or deletions among the genotrophs compared to Pl, the second where Pl and the genotrophs have a large number of SNPs over a short region of the genome. Again, one of the sets of lines is the same as Bethune and the other is very different but each of the variants has a consistent sequence. One of the unanswered questions is the source of these variants since they are not present in an intact form in the progenitor genome. When the corresponding regions of the genomes of other flax accessions and even the wild progenitor of flax, *Linum bienne*, are characterized, the same two alternative structures are seen. The identification of the genes involved in producing these large reproducible modifications of the genome is ongoing.

W310: Flax Genomics

Genomic Analysis Reveals Molecular Mechanisms Underlying Enhanced Alpha-Linolenic Acid Accumulation in Developing Flax Seed

Xue Pan, University of California, Riverside, Riverside, CA and Randall J. Weselake, University of Alberta, Edmonton, AB, Canada

Flax (*Linum usitatissimum* L.), which produces oil containing a high amount of α-linolenic acid (ALA; $18{:}3^{cis\Delta 9,12,15}$), is considered one of the most important plant-based sources of ALA. The goal of our study was to achieve a better understanding of the molecular mechanisms accounting for the high accumulation of ALA in flax oil. Using the flax genome database, we first identified and functional characterized enzymes involved in the final step of triacylglycerol (TAG) biosynthesis, including seven acyl CoA:diacylglycerol acyltransferases (DGATs) and six phospholipid:diacylglycerol acyltransferases (PDATs). Our data provided several lines of evidence in support of flax containing unique PDATs, which can efficiently transfer ALA to TAG. We further performed a comprehensive genome-wide analysis of the *PDAT* gene family across green plants and revealed that the functional divergence of PDAT paralogs is mainly caused by different selection pressures during evolution. Furthermore, our study revealed a second mechanism for enriching the ALA content of TAG, which involves coupling of the DGAT-catalyzed reaction for TAG production to the reverse reaction catalyzed by acyl-CoA:lysophosphatidylcholine acyltransferase (LPCAT). Our *in vivo* and *in vitro* data support the hypothesis that the phosphatidylcholine (PC) de-acylation reaction catalyzed by the LPCAT reverse action can transfer ALAs produced on PC directly into the acyl-CoA pool and make them available for the DGAT-catalyzed reaction for incorporation into TAG. The knowledge obtained from this work will benefit the future development of novel biotechnological strategies to produce value-added seed oils in flax and other oilseed crops.

W311: Flax Genomics

Developing EMS Flax Mutant Lines with Altered SDG Lignan Glucosides

Bourlaye Fofana[1], Kaushik Ghose[1], Jason McCallum[1], Ashok Somalraju[1], Sylvie Cloutier[2], Michael Deyholos[3] and Gordon Rowland[4], (1)Agriculture and Agri-Food Canada, Charlottetown, PE, Canada, (2)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (3)University of British Columbia, Kelowna, BC, Canada, (4)University of Saskatchewan, Saskatoon, SK, Canada

Health benefits for flax seed and its seed coat-containing secoisolariciresinol (SECO) diglucoside (SDG) lignan are rising. SDG is known to be less bioavailable to human intestinal Caco-2 cells than its aglycone SECO, which along with the monoglucoside SMG, are not found naturally *in planta*. Recently, *UGT74S1* was identified and characterized as SDG glucosylation gene. However, whether *UGT74S1* mutants would produce altered SDG profiles *in planta* was unknown. We developed and characterized a flax EMS mutant population by functional genomics approach. Reverse genetics using *UGT74S1* targeted amplicon resequencing of 1996 M2 families identified 94 SNP loci, with 4, 44, 8, and 38 loci in the 5'UTR, exon 1, intron, and exon 2, respectively. The 94 SNP loci covered 138 and 23 SNPs detected as heterozygotes and homozygotes, respectively, for a total of 161 SNP mutations in 138 independent M2 families. Eighteen of the 23 homozygote mutations were in exons. Among those, two were nonsense and thirteen missense, of which two located in the PSGP region of 3 M2 lines. Forward genetics assessing the plant characteristics showed morphological variants, and 69 M2 families carrying UGT74S1 exonic mutations displayed large diversity in SDG profiles, highlighting segregation for the trait. M3 and M4 generations were generated from 30 selected families and genotyped using KASP genotyping assays. Five homozygote M3 lines were identified and of which 2 lines showed null SDG production. The data will be discussed in line with the mutational sensitivity of EMS-induced amino acids towards developing flax lines with highly bioavailable metabolites.

W312: Forage, Feedstocks & Turf

Global Reprogramming of Transcription and Metabolism in Response to Drought, and Genome Wide Association Studies of Drought Related Traits in Medicago

Michael K. Udvardi, The Samuel Roberts Noble Foundation, Ardmore, OK

Periodic drought is a primary limitation on plant growth and yield of crops in many agricultural systems. We have taken two complementary genome-based approaches to identify genes and biological processes that may underpin drought tolerance in the model legume, *Medicago truncatula*, with a view to improving drought tolerance in the important pasture legume, *Medicago sativa* (alfalfa). Transcriptome analysis of roots and shoots from control, mildly, moderately and severely drought-stressed, and re-watered *M. truncatula* and *M. sativa* plants, identified thousands of genes that were altered in expression in response to drought. Expression of many of these genes was tightly coupled to plant water potential (i.e. drought intensity) changes suggesting involvement in drought adaptation. Combining metabolomic data with the transcriptomic data yielded insights into the regulation of metabolic pathways during drought stress. Among the metabolites detected in drought-stressed Medicago plants, proline, myo-inositol and pinitol had striking regulatory profiles indicating involvement in Medicago drought tolerance (Kang et al. 2011, Plant J. 68, 871–889; Zhang et al. 2014, Plant Cell Environ. 37, 2553-76).
In complementary work, we carried out genome wide association studies (GWAS) to identify genes associated with drought adaptation and biomass production in 220 *M. truncatula* ecotypes from the so-called HapMap population (Kang et al. 2015, Plant Cell Environ. 38, 1997–2011). Characterized traits included shoot biomass, maximum leaf size, specific leaf weight, stomatal density, trichome density, and shoot carbon-13 isotope discrimination (δ13C) of well-watered *M. truncatula* plants, and leaf performance under *in vitro* dehydration stress. Reverse-genetics, using a *Tnt1*-insertion mutant population (Cheng et al. 2014, New Phytol. 201, 1065-76) is being used to test predicted roles of specific genes in abiotic stress tolerance.

W313: Forage, Feedstocks & Turf

Genomic and Transcriptomic Analysis of Perennial Ryegrass/Epichloë Endophytes Symbiota

Tim Sawbridge, Department of Economic Development, Jobs, Transport and Resources, Bundoora, Australia; LaTrobe University, Bundoora, Australia

The symbiotic association between perennial ryegrass and *Epichloe* endophytes has been known and studied for decades. The advent of 'next-generation' sequencing has enabled the genome sequencing of both the host grass and the fungal endophytes. Genome sequencing of isolated *Epichloe* fungal endophytes has confirmed the identifies of 4 taxa that colonise perennial ryegrass. Genomic sequencing of strains within taxa have identified a 'core' and 'disposable' complement of the genomes within taxa. Gene loss has been identified in all taxa, however the degree of gene loss is thought to correlate with how recently the 'asexualistion' has occurred within a taxa. Representative strains of all four taxa have been inoculated into breeding germplasm by our commercial partner. This has enabled the transcriptomic analysis of a timecourse of developing symbiota to be undertaken for 3 taxa within the seed of 1 commercial cultivar. This has been achieved through mapping of RNA derived reads to genes from the host grass and the fungal endophytes simultaneously. Results of this analysis covering timepoints from post imbibition to 10 days old seedlings will be presented

W314: Forage, Feedstocks & Turf

Diversification and Use of Bioenergy to Maintain Future Grasslands

Iain Donnison, Aberystwyth University, Aberystwyth University, United Kingdom

W315: Forage, Feedstocks & Turf

Modelling Leaf Growth and Transcriptional Response of Perennial Ryegrass to Drought Stress

Steven Yates[1], Bruno Studer[2], Achim Walter[3], Frank Liebisch[4], Sebastian Nagelmüller[5], Kristina Jonavièienë[6], Gintaras Brazauskas[6] and Norbert Kirchgessner[3], (1)Swiss Federal Institute of Technology, Zuerich, Switzerland, (2)ETH-Zurich, Zurich, Switzerland, (3)ETH Zurich, Zurich, Switzerland, (4)ETH Zurich, Lindau, Switzerland, (5)ETH Zurich, Zuirch, Switzerland, (6)2Institute of Agriculture, Këdainiai, Lithuania

Summer droughts are becoming increasingly common which limit growth in perennial ryegrass. While the growth of leaves has been studied in relation to temperature, the impact of reduced water availability has received little attention. Furthermore the molecular mechanisms regulating growth under water limitation are largely unknown. Here we sought to identify key physiological changes, resulting from drought, and correlate these with transcriptome changes in the meristem. We constructed a model to identify when a plant limits and stops leaf elongation as a consequence of water limitation. This was coupled with transcriptome profiling of the meristem over five days at three times per day, to provide high resolution chronological molecular profiling. By combining both physiological and molecular data, we provide first clues into gene networks controlling growth in perennial ryegrass under abiotic stress. These data provide a rich resource of candidate genes for crop improvement. The results can be used to refine quantitative trait loci regions and identify their underlying causal genes. The relatively rapid screening involved to phenotype growth under water limitation is also applicable to improve yield of perennial ryegrass under summer droughts through selection.

W316: Forage, Feedstocks & Turf

Use of Low-Depth GBS Data for Genomic Prediction Across Different Multi-Parental Pools and Single Plants in Perennial Ryegrass (*Lolium perenne* L.)

Fabio Cericola[1], Luc L. Janss[1], Christian Sig Jensen[2], Morten Greve-Pedersen[2], Ingo Lenk[2], Niels Roulund[2], Dario Fè[3], Stephen Byrne[4], Torben Asp[4] and Just Jensen[1], (1)Molecular Biology and Genetics, Aarhus University, Tjele, Denmark, (2)DLF A/S, Store Heddinge, Denmark, (3)DLF Trifolium, Store Heddinge, Denmark, (4)Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

A cost efficient genotyping strategy is pivotal for introducing genomic prediction (GP) in plant breeding schemes. Allele-proportions estimates derived by genotyping-by-sequencing (GBS) were efficiently used for GP in periennal ryegrass. Increasing the multiplexing of GBS decreases the genotyping price per unit but also reduces the sequencing depth (SD) per marker and an increased number of missing values. This resuls in a decreased accurancy of the allele-proportion estimation. Both leads to biases in the genomic realtionship matrix (GRM) used in gBLUP GS. The present work shows methods to estimate and correct the bias affecting the GRM when SNPs at low depth are used. Moreover three methods for imputing missing values, Mean imputation, k-near neighbour, random forest, were tested.

The usefulness of these methodologies to improve the GP accuracies were shown in simulated data with progressively reduced SD. Furthermore, these findings were applied to real data and allowed to efficently combine samples of perennal ryegrass multi-parental genotype-pools and single plants genotype in different assays at different SD. After imputation and bias correction, increased prediction accuracies were observed for single plants (up to +0.05 for heading date), biparental-pools (up to +0.05 for seed yield) and multiparental-pools (up to +0.07 for seed yield).
Our results provide new methodologies to work with GBS estimated allele-proportions which can be obtained for genotype pools or polyploid crops even at moderate or low sequencing depth. Moreover, the feasibility of predicting performances across multi-parental genotype-pools and single plants gave promising perspectives for implementing GP in perennal ryegrass breeding schemes.

W317: Forage, Feedstocks & Turf

Subterranean Clover(*Trifolium subterraneum* L.) Genomic Resources: Building a Comprhensive Platform for Molecular Breeding

Parwinder Kaur[1], Hideki Hirakawa[2], Philipp E. Bayer[3], Gabriel Keeble-Gagnere[4], Kenta Shirasawa[5], Soichiro Nagano[5], Rudi Appels[4], David Edwards[3], Phillip Nichols[6], Sachiko Isobe[5] and William Erskine[1], (1)Centre for Plant Genetics and Breeding (PGB), University of Western Australia, Perth, WA, Australia, (2)Kazusa DNA Research Institute, Kisarazu, Chiba, Japan, (3)University of Western Australia, Perth, Australia, (4)Murdoch University, Perth, Australia, (5)Kazusa DNA Research Institute, Kisarazu, Japan, (6)Department of Agriculture and Food Western Australia, Perth, Australia

Subterranean clover (*Trifolium subterraneum L.)* is the key forage pasture legume species in Australia and is also an ideal reference species for comparative mapping within the Leguminosae. It is a diploid (2n =16), predominantly inbreeding annual species that can be readily hybridised. It also has a wide range of diversity for both qualitative and quantitative agronomic and morphological characters. A draft genome sequence of size ~480Mb has been described for subterranean clover (based on cv. Daliak). This represents 86% of the estimated 556.6 Mb subterranean clover genome. We used an Illumina next-generation sequencing platform to generate 277.1 Gb (HiSeq) + 23.6 Gb (MiSeq) of genome sequence, which along with sequences derived from a Roche 454 machine (2.72 Gb) enabled us to assemble the pseudo-chromosomes anchored to a high density linkage map based on a custom designed Affymetrix Axiom SNPchip. A total number of 65,295 genes were predicted using Augustus and 18,023 of these were homologous with *Medicago truncatula* (4.0v1).
This reference genome will serve as a genomic resource of subterranean clover and facilitate identification of the genetic basis of agronomically important traits, which are vital for the future genetic improvement of pasture legumes. Having a fully sequenced genome has profound implications for the pre-breeding and molecular marker development of subterranean clover. Presently, a panel of 97 world core collection accessions and 28 elite Australian cultivars, has been extensively phenotyped and genotyped to construct a *haplotype map* (HapMap) which will serve as a powerful tool enabling identification of the genes and genetic variations for important economic and morphological traits.

W318: Forest Tree

Discovery and Transcriptional Dynamics of Small Noncoding RNAs in Source, Transport and Woody Sink Tissues in *Eucalyptus grandis*

Drew Behrens, Alexander Myburg and Eshchar Mizrachi, Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), University of Pretoria, Pretoria, South Africa

Rapid progress is being made in understanding the transcriptional networks and interactions governing xylogenesis and the regulation of lignocellulosic biopolymer synthesis in plants. Despite this, much is still unknown regarding the small noncoding RNAs (sRNAs) that regulate gene expression through complementary sequence targeting, RNA degradation and epigenetic modifications. miRNAs are a well known category of sRNA species, and several highly conserved miRNAs are known to play essential roles in vascular tissue differentiation. The small interfering RNAs (siRNAs) are another more complex and diverse, yet highly pervasive sRNA type in plant genomes, and may indeed contribute significantly to species-specific biology. Here, we report on the discovery and quantification of a comprehensive catalog of *Eucalyptus grandis* sRNAs – including miRNAs and siRNAs – in leaves, secondary phloem and immature xylem, representing major points in carbon sequestration, transport and utilization for lignocellulosic biomass, respectively. Expression analysis of these sRNAs has revealed tissue- specific clusters occuring in vascular and leaf tissue (including high-confidence examples of known miRNA:target interactions). Target prediction and gene set enrichment analysis revealed new regulatory interactions in the xylogenesis transcriptome. In parallel, long noncoding RNA (lncRNA) discovery and expression analysis is being performed to gain insight into potential competition and sequestration of these sRNA species. These putative competitive endogenous (ce)RNAs, their diversity and regulatory functions in xylogenesis remain largely unexplored. This work will allow revision of the multidimensional transcriptional modules underlying xylogenesis in *Eucalyptus*, and may provide insights necessary for the genetic improvement of wood production in this lignocellulosic biomass crop.

W319: Forest Tree

Does Douglas-fir Celebrate the Solstice? Circadian and Circannual Cycles of Gene Expression Variation in Douglas-fir Needles

Peter Dolan[1], Richard Cronn[2], Sanjuro Jogdeo[3], Dee Denver[3], Brad St Clair[2] and Jill Wegrzyn[4], (1)University of Minnesota, Morris, Morris, MN, (2)USDA Forest Service Pacific Northwest Research Station, Corvallis, OR, (3)Oregon State University, Corvallis, OR, (4)University of Connecticut, Storrs, CT

Perennial growth in plants is the product of interdependent cycles of daily and annual environmental stimuli that induce periods of physiological growth and dormancy. In conifers, leaves (“needles”) are a perennial organ that integrates seasonal signals from light, temperature and water availability. To understand the relationship between seasonal stimuli and seasonal responses in conifers, we examined transcriptome changes in Douglas-fir (*Pseudotsuga menziesii*) needles at diurnal and circannual scales. Using mRNA sequencing, we generated 6.1x109 100 bp microreads from 19 trees and constructed a de novo pan- transcriptome reference that includes 162,326 transcripts. Using this reference, we mapped RNA-Seq reads from 166 samples that capture daily, seasonal, and annual variation. Our analysis identifies 15,487 diurnally-cycling transcripts, 4,912 of which show high amplitudes and include homologues to core clock genes from model plants. Analysis also reveals 24,688 annually-cycling transcripts, 11,963 of which show high amplitudes. The timing of maximum gene expression across diurnal and annual periods

shows a bimodal response, with ~50% of transcripts reaching maximum expression +/- 2 hours from sunrise or sunset, and +/- 20 days from the shortest and longest photoperiod. The striking increase in transcription during short photoperiods is unusual because it coincides with dormancy; this may represent an exceptional case of transcriptional “anticipation” for the onset of spring growth. Our results implicate photoperiod as the dominant driver of annual transcription patterns and may be general for temperate zone conifers, making them useful for predicting rhythmic transcription in newly-emerging conifer models.

W320: Forest Tree

A Transcriptomics Approach to the Development of Predictive Molecular Markers for Tolerance to Tree Diseases: Ash Dieback

Andrea L. Harper[1], Lea V. McKinney[2], Lene R. Nielsen[2], Lenka Havlickova[1], Yi Li[1], Martin Trick[3], Fiona Fraser[3], Lihong Wang[1], Alison Fellgett[1], Elizabeth S. A. Sollars[4], Sophie H. Janacek[5], J. Allan Downie[3], Richard Buggs[4], Erik D. Kjær[2] and Ian Bancroft[1], (1)University of York, York, United Kingdom, (2)University of Copenhagen, Copenhagen, Denmark, (3)John Innes Centre, Norwich, United Kingdom, (4)Queen Mary University of London, London, United Kingdom, (5)The Genome Analysis Centre, Norwich, United Kingdom

Elucidating mechanisms and exploiting genetic tolerance to pests and diseases are major challenges in tree species. Ash dieback disease, caused by the fungal pathogen *Hymenoscyphus fraxineus*, is killing *Fraxinus excelsior* throughout Europe. To identify the molecular basis of the tolerance shown by a very small proportion of trees, with the aim of developing predictive molecular markers to assist breeding, we deployed for the first time in a tree species the technology of Associative Transcriptomics. This method, first developed in the annual crop species *Brassica napus* (Harper et al. Nature Biotechnology 30:798-802, 2012), combines rapid marker discovery for both gene sequence and gene expression variation with association genetic and systems biology approaches. We established functional genotypes for a training diversity panel of 186 trees with varying levels of disease susceptibility, comprising 174,470 Single Nucleotide Polymorphism (SNP) markers and 32,441 Gene Expression Markers (GEMs). We were able to find highly significant ($P < 10^{-8}$) associations between the extent of canopy damage and both SNP markers and GEM markers. Predictive capability within the training panel was confirmed using a “take one out” assessment, so the top 3 markers were used to successfully predict canopy damage ($R^2 = 0.24$) in a test dataset of 58 additional, unrelated accessions based on differential gene expression levels in freshly flushed leaves, with prediction of low damage (highly tolerant) individuals being remarkably successful.

W321: Forest Tree

A Comprehensive Study of the Sugar Pine (*Pinus lambertiana*) Transcriptome Implemented through Diverse Next-Generation Sequencing Approaches

Pedro J Martínez-García[1], Daniel Gonzalez-Ibeas[2], Randi Famula[1], Annette Delfino-Mix[3], Kristian Stevens[4], Jeffrey Puryear[5], Charles H. Langley[4], Carol Loopstra[5], David Neale[1] and Jill Wegrzyn[2], (1)Dept. Plant Sciences University of California Davis, Davis, CA, (2)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (3) Institute of Forest Genetics, USDA Forest Service, Placerville, CA, (4)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (5)Dept. of Ecosystem Science and Management, Texas A&M University, College Station, TX

The assembly, annotation, and characterization of the sugar pine (*Pinus lambertiana* Dougl.) transcriptome represents an opportunity to study the genetic mechanisms underlying resistance to the invasive white pine blister rust (*Cronartium ribicola*) as well as responses to other abiotic stresses. The assembled transcripts also provide a resource to improve the genome assembly. We selected a diverse set of tissues allowing the first comprehensive evaluation of the sugar pine gene space. We have combined short read sequencing technologies (Illumina MiSeq and HiSeq) with the relatively new Pacific Biosciences Iso-Seq approach. From the 2.5 billion and 1.6 million Illumina and PacBio (46 SMRT cells) reads, 33,720 unigenes were *de novo* assembled. Comparison of sequencing technologies revealed improved coverage with Illumina HiSeq reads and better splice variant detection with PacBio Iso-Seq reads. The genes identified as unique to each library ranges from 199 transcripts (basket seedling) to 3,482 transcripts (female cones). In total, 10,026 transcripts were shared by all libraries. Genes differentially expressed in response to these provided insight on abiotic and biotic stress responses. To analyze orthologous sequences, we compared the translated sequences against 19 plant species, identifying 7,229 transcripts that clustered uniquely among the conifers. We have generated here a high quality transcriptome from one WPBR susceptible and one WPBR resistant sugar pine individual. Through the comprehensive tissue sampling and the depth of the sequencing achieved, detailed information on disease resistance can be further examined.

W322: Forest Tree

Identification of Gene Sets with Potential Roles in Biotic Stress Defenses in Western Redcedar by Comparative RNA-Seq Analysis

Jim Mattsson, Simon Fraser University, Biological Sciences, Burnaby, BC, Canada

Western redcedar *(Thuja plicata)* trees can live for more than 1000 years, indicating exceptionally strong biotic resistance. An example of that is the naturally rot resistant heartwood, popular for various outdoor applications. Nevertheless, reforestation is inefficient and expensive due to extensive deer and elk browsing, and second-growth trees are frequently afflicted by extensive heart rot, jeopardizing the reputation of durable cedar lumber. We used RNA-seq to identify candidate genes and chemical biosynthesis pathways contributing to biotic resistance in both foliage and heartwood-forming tissues. First, a comparison of transcriptomes from wildtype foliage and a natural variant lacking resin glands yielded +500 genes, including complete biosynthetic pathways of foliar terpenoids, compounds that are linked to browsing resistance. Functional assays identified sabinene synthase and hydroxylases, key enzymes in the biosynthesis of the deer repellant a-thujone. Secondly, a comparison of transcriptomes from the sapwood to heartwood transition zone with other wood fractions provided candidates for complete biosynthetic pathways for tropolones and lignans, compounds that are linked to heartwood rot resistance, and surprisingly also pathways for flavonoid biosynthesis, compounds that hitherto have not been linked to rot resistance in this species. Finally, the bark-forming region also expressed genes in lignan, terpenoid and flavonoid biosynthesis, suggesting that the bark and heartwood have similar biotic defenses, but also expression

indicating production of bark-specific toxins such as phenazines. Identified genes are now being explored for use in multi-trait Genomic Selection in this economically important species.

W323: Forest Tree

A Comprehensive lincRNA Analysis: From Conifers to Trees

Nicolas Delhomme[1], Bastian Schiffthaler[1], Katja Stojkovic[2], Yao-Cheng Lin[3] and Nathaniel R. Street[1], (1)Umeå Plant Science Centre, Umeå University, Umeå, Sweden, (2)Umeå Plant Science Center, Umeå, Sweden, (3)VIB - Ghent University, Zwijnaarde, Belgium

We have produced an updated annotation of the Norway spruce genome on the basis of an *in silico* normalised set of RNA-Seq data obtained from 1,529 samples and comprising 15.5 billion paired-end Illumina HiSeq reads complemented by 18Mbp of PacBio cDNA data (3.2M sequences). In addition to augmenting and refining the previous protein coding gene annotation, here we focus on the addition of long intergenic non-coding RNA (lincRNA) and micro RNA (miRNA) genes.

In addition to non-coding loci, our analyses also identified protein coding genes that had been missed by the initial genome annotation and enabled us to update the annotation of existing gene models. In particular, splice variant information, as supported by PacBio sequencing reads, has been added to the current annotation and previously fragmented gene models have been merged by scaffolding disjoint genomic scaffolds on the basis of transcript evidence. Using this refined annotation, a targeted analysis of the lincRNAs enabled their classification as i) deeply conserved, ii) conserved in seed plants iii) gymnosperm/conifer specific.

Concurrently, complementary analyses were performed as part of the aspen genome project and the results of a comparative analysis of the lincRNAs conserved in both Norway spruce and Eurasian aspen enabled us to identify conserved and diverged expression profiles.

At present, we are delving further into the expression results with the aim to functionally annotate the lincRNA genes, by developing a co-expression network analyses based GO annotation.

W324: Forest Tree

Gene Expression Profiles of Low Temperature Related Genes in Cold Resistant and Sensitive Black Poplar (*Populus nigra* L.) Clones

Evrim Zeybek, Kubilay Yildirim and **Zeki Kaya**, Department of Biological Sciences, Middle East Technical University, Ankara, Turkey

Low temperature is one of the major environmental constraint which affect the survival of woody plants. Black poplar (*Populus nigra*) has a broad diversity of proteomic, biochemical and physiological mechanisms to cope with low temperature effects. Low temperature not only induces different types of responses, but also, it triggers many pathways that bring about differential gene expression in black poplar. In this study, four important time/temperature points were selected to identify the low temperature induced genes in cold resistant and cold sensitive black poplar clones by using microarray techniques. The stem sections were collected in September 2011 (19.32°C) as control, in November 2011 (before cold: 1.61°C), in February 2012 (during cold: -2.53°C) and in April 2012 (after cold: 13.76°C). The differential expression of low temperatures related genes were examined in these four period. A statistical threshold for the microarray data analysis detected a total of 3983 genes significantly expressed (Fold Change (FC)≥5, p value≤0.01) in cold resistant clones. Among these genes, 1649 were significantly expressed on February, about 650 of these were up regulated while over 1000 of them were down regulated in cold resistant clone. In cold sensitive clone, 3732 genes were differentially expressed and 1817 of them with similar pattern as in cold resistant clone were significantly expressed on February 2012. In contrast to differentially altered genes in the resistant genotype, low temperature had more impact on the transcriptome of the sensitive genotype. In November, carbohydrate metabolism seems to be important for black poplar clones for preparing themselves to winter conditions while in February, carbohydrate metabolism and transport activity related genes were expressed. On the other hand, genes related to resonse to cold were highly expressive in cold sensitive clone. Information generated from this study provides new data on list of possible candidate genes involved in cold response in black poplars that could be highly useful for future studies dealing with mechanisms of chilling tolerance in woody plants.

W325: Forest Tree

Toward Higher Resolution Co-Expression Networks of Wood Developing Tissues through Data Integration

Matthew S. Zinkgraf[1], Lijun Liu[1], Trevor Ramsay[2], Andrew T. Groover[1] and Vladimir Filkov[2], (1)US Forest Service, Davis, CA, (2)UC Davis, Davis, CA

Trees grow in a diverse array of habitats and must integrate environmental and developmental signals during the formation of woody tissues. Increasing evidence suggests that these signals are routed through complex transcriptional networks where they are precisely aggregated to elicit appropriate gene expression responses. To resolve such signal-aggregating regulatory pathways involved in wood formation, approaches that consider multiple experiments at the same time are needed. In this talk, we present a data integration analysis of diverse genomic datasets from *Populus* aimed at more precise understanding of transcriptional and regulatory responses during the wood formation process. Using multiple transcriptome profiling (RNA-seq) experiments, we generated higher-resolution gene co-expression networks, and identified both gene modules conserved across multiple datasets and those limited to specific datasets. Modules were identified that have significant correlations with experiment specific phenotypes, and a meta analysis of these datasets reveals that a subset of these same gene modules were highly conserved across multiple datasets. Furthermore, integration of transcription factor binding experiments (ChIP-seq) reveals that conserved gene modules were enriched for binding from three transcription factors (BELLRINGER, ARBORKNOX 1 and ARBORKNOX 2) that play fundamental roles in vascular cambium regulation. Together, this integrated approach increases the specificity of co-expression models and allows for the identification of conserved regulatory interactions, and suggests that a limited number of genes may be broadly involved in integrating environmental and developmental signals during wood formation.

W326: Forest Tree

A Genome-Wide Association Study Including Common and Rare Genetic Variants Reveals Putative Regulators of Bioenergy Traits in Eastern Cottonwood

Annette M. Fahrenkrog[1], Leandro Gomide Neves[1], Marcio F Resende Jr.[2], Christopher Dervinis[1], Brad Barbazuk[1] and Matias Kirst[1], (1)University of Florida, Gainesville, FL, (2)RAPiD Genomics LLC, Gainesville, FL

The growing bioenergy industry requires an increase in availability of plant feedstocks with high biomass to biofuel conversion efficiency. Poplars are a suitable feedstock for bioenergy production, but cultivars specifically tailored for the biofuel industry are still under development. Understanding the genetic control of bioenergy-related traits can address this gap by accelerating poplar breeding. Genome-wide association studies (GWASs) have been successful in identifying common genetic variants that regulate commercially important traits in crop plants. Although these studies have revealed part of the genetic component that explains trait variance, much of the heritability remains missing. Rare variants may account for part of the missing heritability, especially in species with high genetic diversity such as forest trees. Here we report the first GWAS conducted in *Populus deltoides*, one of the main poplar species used for breeding, analyzing association of common and rare genetic variants with bioenergy traits. A population of 500 unrelated individuals was genotyped by resequencing 18,153 genes, followed by SNP identification. Single- and multiple-marker association analysis was performed, identifying significantly associated genes with eight biomass growth and wood composition traits phenotyped in the population. Including rare variants in this GWAS increased the number of associations identified by 62%. These results highlight the importance of including low frequency variants in GWAS of highly diverse tree species. This was made possible by utilizing flexible genotyping methods such the sequence-capture/next-generation sequencing approaches, and by applying statistical methods designed to detect trait associations with rare variants.

W327: Forest Tree

Genotypic-Phenotypic Variation and Marker-Based Heritability Estimates of a Shrub Willow (*Salix purpurea*) Association Population

Fred E. Gouker[1], Ran Zhou[2], Luke Evans[2], Stephen DiFazio[2], Ben Bubner[3], Matthias Zander[4] and Lawrence Smart[1], (1)Cornell University, Geneva, NY, (2)West Virginia University, Morgantown, WV, (3)Thünen-Institute of Forest Genetics, Waldsieversdorf, Germany, (4)Humboldt University of Berlin, Berlin, Germany

Salix spp. and hybrids (shrub willow) have been bred as dedicated bioenergy crops, however there is still untapped potential for genetic improvement. One way to accelerate and improve upon the breeding and selection is to utilize the large amount of genetic variation present within the genus through the collection of phenotypic datasets, high-density genotyping-by-sequencing and genetic mapping studies. *Salix purpurea* is a reference species for breeding shrub willow bioenergy crops in North America. We assembled a genetically diverse germplasm collection of 112 accessions and obtained high-quality data for 21 biomass, morphological, phenological, physiological and wood compositional traits from three replicated field trials across two years. With a filtered set of 25,566 GBS markers, marker-based estimations of narrow sense heritability of all traits were calculated and ranged between 0.24-0.89 while estimates based on genotypic means ranged from 0.29-0.87. Genotypic and phenotypic correlations as well as multiple linear regression using step-wise selection of all traits were highly correlated with final biomass yield at harvest with the highest correlation of r=0.78. The North American collection of *S. purpurea* was also expanded through the addition of 165 accessions from Europe which revealed population stratification and subdivisions using GBS SNP markers. Using STRUCTURE and DAPC to characterize population genetic structure, results showed five to six subpopulations, respectively. The expansion of this analysis with accessions from the native range of *S. purpurea* provides an opportunity for enhanced studies of phenotype-genotype associations and the genetic bottleneck that occurred through naturalization of *Salix purpurea* in North America.

W328: Forest Tree

A New Look at Population Recombination and Linkage Disequilibrium in Forest Trees from Genome-Wide SNP Data in *Eucalyptus* and its Relevance to Molecular Breeding

Orzenil Bonfim Silva-Junior[1,2] and **Dario Grattapaglia**[2,3], (1)Bioinformatics Lab - EMBRAPA Genetic Resources & Biotechnology, Brasilia, Brazil, (2)Graduate Program in Genomic Sciences - Catholic University of Brasilia, Brasilia, Brazil, (3)Plant Genetics Lab - EMBRAPA Genetic Resources & Biotechnology, Brasília, Brazil

Understanding the genome-wide patterns of recombination provides important insights into the shaping of the species' genetic history and on the extent of linkage disequilibrium (LD), which in turn determines our ability to dissect quantitative traits and predict complex phenotypes. We characterized the population-scaled recombination rate (ϱ) and LD in *E. grandis* using 21,517 polymorphic genome-wide Infinium SNPs and nearly 13 million SNPs from whole-genome pooled resequencing of 36 individuals. At the genome-wide scale, LD decayed within ≈4-6 kb, considerably slower than previously reported from candidate gene studies, but showing variation from absence to complete LD up to 50-kb. A sharp decrease in the estimate of ϱ was seen when going from short (0-100 bp) to genome-wide (0-50 kb) inter-SNP distances, highlighting the dependence of ϱ on the scale of observation, while further corroborating that LD estimates from candidate gene studies cannot provide accurate expectations of genome-wide LD. Our results challenge the heretofore-established consensus of rapid decay of LD in outcrossed forest tree genomes, converging with other recent genome-wide studies in Poplar. The more extensive LD found has considerable practical importance to molecular breeding. While a genome-wide usable LD within ≈4-6 kb in natural populations warrants a positive outlook on the prospects of GWAS from the detection standpoint, it complicates pinpointing causative QTNs, if such a feat will ever be accomplished. On the other hand, with extensive LD in closed breeding populations, low-density genotyping platforms will provide abundant power to capture linked effects for whole-genome prediction at tree breeders' accessible costs.

W329: Forest Tree

Comparing Sequence-Capture and SNP-Array Genotyping Methods for Development of Genomic Selection Prediction Models in *Eucalyptus*

Rodrigo Furtado dos Santos[1], Braulio Fabiano Xavier de Moraes[2], Alexandre Missiaggia[3], Aurélio M Aguiar[3], Bruno M. Lima[4], Donizete da C. Dias[3], Gabriel D. P. S. Rezende[3], Flávia M. A. Gonçalves[5], Marcio Resende[6], Patricio R. Munoz[7] and Matias Kirst[1],

(1)University of Florida, Gainesville, FL, (2)Federal University of Lavras, Lavras, Brazil, (3)Fibria Celulose S.A., São Paulo, SP, Brazil, (4)FIBRIA S.A. Technology Center, Jacarei, Brazil, (5)Federal University of Lavras, Lavras, MG, Brazil, (6)RAPiD Genomics LLC, Gainesville, FL, (7)University of Florida, Agronomy Department, Gainesville, FL

High-throughput and low-cost methods of genotyping are critical for the successful application of genomic selection (GS) approaches. Until recently, GS studies in trees have predominantly relied on SNP arrays as the source of genotypic data. However, this technology has a high entry cost that can prevent the use of GS in many tree breeding programs. Also, tree breeding programs typically employ locally adapted species and populations, hampering the worldwide transferability of a standard genotyping platform. Advances in next-generation sequencing and methods of genome complexity reduction, such as sequence-capture (*a.k.a.* exome capture), allow the development of flexible genotyping platforms that can be tailored to specific species and populations with a low entry cost. However, the suitability of sequence-capture approaches to genotyping and development of GS prediction models has not been evaluated, or compared to models developed from SNP arrays. Here we evaluate the impact of sequence-capture and array-based genotyping methodologies on the development of GS prediction models for a *Eucalyptus* breeding population from Fibria Celulose S.A., composed of 739 trees phenotyped for 13 wood quality and growth traits. The descriptive analysis of the datasets (pattern of linkage disequilibrium, minor allele frequency and missing data content) was used to compare both methods. RR-Blup and BayesB prediction methods were developed using data from both genotypic datasets, and predictive ability estimated using cross validation was employed to evaluate the performance of GS models. Differences in linkage disequilibrium patterns, minor allele frequency, missing data and marker distribution were detected among datasets derived from sequence-capture and SNP arrays. However, models RR-Blup and BayesB resulted in similar predictive abilities and demonstrated that both genotyping methodologies are equivalents for the traits evaluated.

W330: Forest Tree

Alternative Approaches to Loblolly Pine Breeding Value Predictions

Adam R Festa, R.W. Whetten and Lilian Matallana, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC

The objective of this research is to assess the relative contribution of separate covariance matrices to predict full-sib progeny performance from crosses among a small number of parents, by modeling gene structure and gene expression in cross-validation studies. Questions we intend to address include: (1) can we obtain reasonably reproducible results from triplicate samples of seedlings from OP, PMX, or CP families with respect to estimating family-mean levels of gene expression for a set of parents; (2) can we identify methods for combining those family-mean estimates of gene expression levels into covariance estimates for pairwise-combinations of parents that show utility in cross-validation studies for modeling phenotypic variation, and (3) do covariance matrices based on coding sequence SNP variation, gene expression level variation, or pedigree-based estimates of allele sharing have independent value for modeling phenotypic variation, or are they redundant so that one approach has the same information present in the other two? To answer these questions, we have completed RNA-seq on 43 different families using biological replication of pools of whole seedlings harvested by family at age 3-months. Family-mean gene expression patterns were used with phenotypes from age 6 progeny tests to identify genes or clusters of genes deemed as having a relationship with the trait. By utilizing covariance matrices of these genes we have produced cross validation predictions with R^2 values of .5-.59. To assess bias in our approach, we are sequencing an additional set of 31 families for further analysis.

W331: Forest Tree

High-Resolution Mapping of Biomass-Related Traits in Shrub Willow (*Salix purpurea* L.)

Craig H. Carlson[1], Fred E. Gouker[1], Stephen DiFazio[2], Ran Zhou[2] and Lawrence Smart[1], (1)Cornell University, Geneva, NY, (2)West Virginia University, Morgantown, WV

Shrub willow (*Salix* spp.) is a vigorous woody perennial feedstock bred for dedicated biofuel production. Recent genomic advances have provided the biomass feedstock community with new tools to improve traits related to biomass yield and wood chemical composition. Large mapping populations have been generated in order to locate QTL associated with these traits. In 2014, 497 progeny from an F_2 *S. purpurea* family were planted in a randomized complete block design at a location of uniform grade and soil type at the New York Agricultural Experiment Station, Geneva, NY. A total of 40 phenotypes and over 100,000 data points were measured during the establishment and first year post-coppice seasons. Family parents, grandparents, and progeny were genotyped via genotyping-by-sequencing (GBS) using the restriction enzymes, *ApeK*I and *EcoT22*I. Variant discovery was performed in the Tassel v3 GBS pipeline using the *S. purpurea* v1.0 genome (F_2 grandparent) as a reference. All SNPs were quality filtered and duplicate sites collapsed. Any SNPs deviating from expected segregation were removed. Of the initial SNPs, approximately 5,000 were considered informative. Linkage groups were constructed and markers ordered in MSTMap then imported into R/qtl for downstream analyses. Significant QTL were identified for numerous growth traits as well as those for insect and disease resistance. For instance, a major QTL for stem height (LOD=19.3, *pve*=19.03%) resides within a QTL cluster (1-3 cM) for multiple biomass-related traits. Beyond proof-of-concept, this work aims to provide tools for early selection of traits important for biofuel production in shrub willow bioenergy crops.

W332: Forest Tree

Exome Genotyping and Association Genetics of Environmental Adaptation and Stress Mitigation Traits in a Clonally Tested Loblolly Pine (*Pinus taeda* L.) Population

Mengmeng Lu[1], Konstantin V. Krutovsky[1,2], C. Dana Nelson[3], Tomasz Koralewsky[4], Tom D. Bryam[4] and Carol Loopstra[1], (1)Texas A&M University, College Station, TX, (2)Georg-August-University of Göttingen, Göttingen, Germany, (3)Southern Institute of Forest Genetics, US Forest Service SRS, Saucier, MS, (4)Texas A&M Forest Service/Texas A&M University, College Station, TX

Loblolly pine, *Pinus taeda* L., is the most widely planted and commercially important tree species in the southeastern U.S. To increase the number of known single nucleotide polymorphisms (SNPs) and functional markers available for research and tree breeding, we used genotyping by sequencing for targeted exome regions. The exons were captured in 375 trees using NimbleGen probes and then sequenced using the Illumina HiSeq 2500 platform. Hybridization oligonucleotide probes were designed for 199,723 exons (≈49Mb) partitioned from the loblolly pine genome

annotation v2 (PineRefSeq). The bioinformatics analyses demonstrated 90.2% of the targeted exons were covered by the probes. The capture efficiency analyses showed that an average of 67.2% of the captured bases were in the target regions and more than 70% of the target bases had at least 10X sequencing depth. A total of 972,720 SNPs were acquired after filtering. Among them, 52.8% were located in coding regions and 5.3% were located in 5' or 3' untranslated regions. Mean heterozygosity (0.32±0.01) and nucleotide diversity (0.27±0.13, on a per site basis) were high within this population. The zygotic linkage disequilibrium (LD) coefficient (r^2) decayed to half within 55 bp, to r^2=0.1 within 192 bp, and r^2=0.05 within 451 bp. The *fastStructure* analysis using unlinked SNPs demonstrated two distinct clusters representing west and east parts of the loblolly pine area. We will describe association tests that are being conducted to discover markers and genome regions associated with phenotypic traits including height, specific leaf area and carbon isotope discrimination.

W333: Forest Tree

Genomic Prediction and Linkage Disequilibrium in *Eucalyptus benthamii* and *Eucalyptus pellita* using a 60K SNPs Chip (EUChip60K)

Bárbara S. F. Müller[1], Leandro G Neves[2], Marcio Resende[2], Annette M. Fahrenkrog[3], Janeo Almeida Filho[3], Patricio R. Munoz[4], Matias Kirst[3] and Dario Grattapaglia[5], (1)University of Brasília - EMBRAPA, Brasília, Brazil, (2)RAPiD Genomics LLC, Gainesville, FL, (3)University of Florida, Gainesville, FL, (4)University of Florida, Agronomy Department, Gainesville, FL, (5)Plant Genetics Lab - EMBRAPA Genetic Resources & Biotechnology, Brasília, Brazil

Eucalyptus benthamii (BEN) and *Eucalyptus pellita* (PEL) are species of increasing commercial interest due to their cold and drought tolerance, respectively. Genomic Selection (GS) is a promising approach to accelerate the breeding cycle and develop elite trees of these unexploited species. We report the development of genomic prediction models for growth traits and estimates of Linkage Disequilibrium (LD) decay in two breeding populations of BEN (*Ne*=53, *n*=505) and PEL (*Ne*=35, *n*=732), using SNPs genotyped with an Ilumina Infinium chip. Prediction models were initially built with 13,787 and 19,506 high quality polymorphic SNPs. Using progressively smaller SNP datasets (10,000 to 100), the results suggests that models with ~5,000 SNPs are equivalent in predictive ability (r_{gy}) to the full model for all traits and species. The r_{gy} inferred by Bayesian methods (BRR, BayesB, BayesA, BayesCπ, BL) reached similar estimates, varying from 0.14 for volume in BEN to 0.44 for DBH in PEL. The lower values of r_{gy} for BEN may be explained by the restricted occurrence of this species in its natural range and the limited genetic diversity sampled by this breeding population. The average genome-wide LD (r^2) dropped below 0.2 within 15.6 and 70.6 Kbp for BEN and PEL, respectively. After correcting for relatedness and structure (r^2VS) LD decayed faster, at 7.7 and 25.5 Kbp, confirming the strong effect of genetic relationship expected in these populations. This study sets the stage for the application of high-density SNPs and GS in two specialty eucalypts, to characterize their populations and advance breeding.

W334: Forest Tree

TreeGenes: Enabling Visualization and Analysis in Forest Tree Genomics

Emily Grau[1], Steven A Demurjian Jr[1], Hans Vasquez-Gross[2], Damian Gessler[3], David Neale[4] and Jill Wegrzyn[5], (1)University of Connecticut, Storrs, CT, (2)University of California Davis, Davis, CA, (3)Semantic Options, Santa Fe, NM, (4)Dept. Plant Sciences University of California Davis, Davis, CA, (5)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT

Successfully identifying genes and alleles associated with traits of interest in forest trees requires robust integration of genotype, phenotype, and environmental data. Feasibility of these association studies depends on next-generation sequence data, ontologically sound phenotypes, and geo-referenced measures. The TreeGenes project provides custom informatics tools and databases to manage the flood of information resulting from high-throughput genomics projects in forest trees, from sample collection to downstream analysis. We will discuss tools available through TreeGenes to access data from over 1700 forest tree species.

The TreeGenes database is designed to store and integrate data from public repositories as well as information that is not typically curated or submitted. In specific, we are actively collecting and curating metadata resulting from genotype:phenotype and genotype:environment association studies. The database is accessible through a variety of interfaces that deliver taxonomy views, literature searches, transcriptome annotations, genetic maps (CMap), genome browsers (Gbrowse), and variant annotations.

TreeGenes hosts custom-developed tools that allow users to: track and manage high-throughput projects through the Forest Tree Genetic Stock Center, query and download bulk datasets through DiversiTree, and perform analysis of data on HPC resources using CartograTree. The upcoming transition to the more standardized Tripal database solution will allow TreeGenes to integrate with other valuable repositories, including the Hardwood Genomics Web and the Genome Database for Rosaceae. The in-development Tripal Galaxy module will allow TreeGenes to offer an expanded set of analytical tools and flexible computing resources to users.

W335: Forest Tree

Genomic Selection for New Zealand Forestry

Natalie Graham[1], Emily J. Telfer[1], Phillip Wilcox[2], Yongjun Li[1], Mari Suontama[1], Mitchell Murray[1,3], Lucy Macdonald[1], Wei-Young Wang[4] and Heidi Dungey[1], (1)Scion, Rotorua, New Zealand, (2)University of Otago, Dunedin, New Zealand, (3)University of Waikato, Hamilton, New Zealand, (4)PF Olsen Limited, Rotorua, New Zealand

Radiata pine is the world's most planted exotic conifer with over four million hectares worldwide, and constitutes 90% of the plantation forest estate in New Zealand. Currently the breeding programme is up to the third generation, suggesting that the genetic potential of this highly heterozygous outcrossing species is still largely untapped. Scion is assisting the Radiata Pine Breeding Company to develop Genomic Selection to accelerate the breeding programme and deliver genetic gains in a much shorter time frame than phenotypic selection. Leveraging off an existing transcriptomic resource, we have developed an exome-capture probe panel with assistance from Rapid Genomics LLC. We present results from the optimisation of this panel and the design of robust filtering pipelines, and report on the estimated genome coverage and the number of SNPs generated using this panel.

The remaining 10% of New Zealand's commercial forest estate is a combination of other exotic softwoods and exotic hardwoods. Within the hardwoods, the *Eucalyptus* spp. are of particular interest. *Eucalyptus nitens* shows promise as a vigorous species for planted forests, and is well-

suited to many New Zealand sites. *Eucalyptus regnans* and *Eucalyptus fastigata* are also popular choices for both pulp and carbon sequestration. The existing multi-species *Eucalyptus* SNP chip EuCHIP60K is facilitating the introduction of genomic selection into these eucalypt breeding programmes. We have successfully genotyped *E. nitens* with the EuCHIP60K and present here results from additional testing of other species including *E. regnans* and *E. fastigata.*

W336: Forest Tree

Genetic Containment of Forest Trees by RNAi Suppression of LEAFY

Amy Leigh Klocko[1], Amy Brunner[2], Jian Huang[3], Richard Meilan[4], Haiwei Lu[1], Cathleen Ma[1], Alice Morel[1], Dazhong Zhao[3], Kori Ault[1], Michael Dow[1], Olga Shevchenko[1] and Steven Strauss[1], (1)Oregon State University, Corvallis, OR, (2)Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, (3)University of Wisconsin-Milwaukee, Milwaukee, WI, (4)Purdue University, West Lafayette, IN

Gene flow into wild and feral populations of forest trees present a significant barrier to field studies and commercial use of exotic and recombinant DNA modified trees. Both male and female sexual reproduction are significant concerns in most species. To provide bisexual containment, we used RNA-interference (RNAi) to suppress the poplar ortholog of the *LEAFY* (*LFY)* gene, which is essential for development of floral primordia in male and female sexual organs. We transformed this construct into male and female clones of poplar; here we present results from early-flowering female clone 6K10 (*Populus alba*). We obtained 15 independent transformed events in clone 6K10 and examined their floral and vegetative traits over 4 seasons of growth in an APHIS-approved field trial. Rooted trees were planted in 2011 and began flowering in 2014. Floral phenotypes were initially assessed through indoor flushing of dormant floral buds followed by observation of field-opened buds. We found that suppression of the poplar orthologs of *LFY* gave rise to complete and stable sexual sterility in the field. Of the 15 RNAi-LFY events, 2 had extremely small inflorescences that lacked functional sexual organs and had reduced expression of *LFY* in developing floral buds. The floral phenotype was repeated over two growing seasons, and the trees showed normal survival, seasonal dormancy, vegetative morphology, and growth rate. Suppression or mutation of this gene should greatly facilitate field research, regulatory approval, and public acceptance of exotic and recombinant DNA modified forest trees.

W337: Forest Tree

Asexual Gene Drive in Populus? Results from CRISPR/Cas9 Mutagenesis of Floral Genes for Genetic Containment

Estefania Elorriaga, Amy Leigh Klocko, Cathleen Ma and Steven Strauss, Oregon State University, Corvallis, OR

Gene flow from genetically engineered (GE) trees into feral or wild populations are significant obstacles to their use as a result of regulatory, public perception, and ecological concerns. Loss-of-function mutations in a number of floral transcription factor genes can lead to sterility in diverse plant species, however such mutations are rare and generally recessive, thus are very difficult to induce via conventional tree breeding. The recently rediscovered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated Cas system has proven to be a powerful directed-mutagenesis tool in many species, including trees. The CRISPR-Cas induced mutations appear to be highly predictable and stable, and reversion should be extremely rare or impossible (e.g., when there are deletions of essential parts of coding regions). We are testing the mutation efficiency of six nucleases targeting two essential floral genes in *Populus*. The targets are the poplar orthologs *LEAFY* and *AGAMOUS*—well studied genes essential for both male- and female-fertility. The nucleases have been stably transformed into hundreds of independent events that we are now analyzing for mutation rate and type. Sequencing results from 300 events indicate that the rate of identical mutation in both *LEAFY* alleles is ~32% and in both *AGAMOUS* alleles is ~27%. This high rate of identical biallelic mutations may be due to CRISPR-Cas-mediated gene drive (allelic conversion) during initial mutagenesis or mitotic growth. Results from study of the full population will be presented.

W338: Forest Tree

Genome Sequencing, Comparative Genomics and Population Analyses of Aspen Species

Yao-Cheng Lin[1], Nicolas Delhomme[2], Jing Wang[3], **Bastian Schiffthaler**[2], Manfred G. Grabherr[4], Neda Zamani[4], Marc Höppner[4], Chanaka Mannapperuma[2], Niklas Mähler[5], David Sundell[2], Yves Van de Peer[1], Torgeir Hvidsten[5], Stefan Jansson[2], Par K. Ingvarsson[3] and Nathaniel R. Street[2], (1)VIB - Ghent University, Zwijnaarde, Belgium, (2)Umeå Plant Science Centre, Umeå University, Umeå, Sweden, (3)Department of Ecology and Environmental Sciences - Umeå Universitet, Umeå, Sweden, (4)Medical Biochemistry & Microbiology, Uppsala University, Uppsala, Sweden, (5)Norwegian University of Life Sciences, Ås, Ås, Norway

Populus tremula and *P. tremuloides* are keystone species used as model systems for forest tree research and provide an ideal system for exploring the genetic architecture of complex phenotypes and studies of local adaptation. Genomics analyses in these species currently rely on the *P. trichocarpa* reference genome, which is limiting due to poor sequence alignment rates. We therefore sequenced two aspen genomes in addition to re-sequencing 24 *P. tremula* and 22 *P. tremuloides* individuals. Additional aspen species (*P. grandidentata* and *P. davidiana)* have also been sequenced. A *P. tremula* intra-specific genetic map using a new F_1 population is being generated.

We produced NGS based *de novo* genome assemblies that were *ab initio* annotated, then supplemented using RNA-Seq transcript profiling, enabling identification of novel, aspen specific genes. Genome assembly was challenging as a result of extensive heterozygosity, however, the gene space is comprehensively represented. Strategies to improve the assembly were investigated, demonstrating the potential for genome assembly using SMRT sequencing data.

This resource enabled comparative analysis to *P. trichocarpa* and among aspen species. Whole genome alignment and population level genotype data identified regions in aspen that have rapidly diverged from the reference *P. trichocarpa* genome. Chloroplast and nuclear genome analyses reconstructed the population history and phylogenetic trees for these species. We are now performing genome wide association studies and extensive RNA-Seq projects investigating controls of wood and leaf development, and exploring sexual dimorphism of gene expression and secondary metabolite biosynthesis.

The resources have been integrated in PlantGenIE.org.

W339: Forest Tree

Multilevel Modelling of Lignocellulosic Biomass Accumulation in Eucalyptus

Eshchar Mizrachi[1], Desre' Pinard[1], ACF Gutiérrez[2], Lieven Verbeke[2], Colan G. Balkwill[1], Drew Behrens[1], Elodie Ekoka[1], Steven G Hussey[1], Yves Van de Peer[3], Kathleen Marchal[3] and Alexander A. Myburg[1], (1)Department of Genetics, Forestry and Agricultural Research Institute and Genomics Research Institute, University of Pretoria, Pretoria, South Africa, (2)Dept. of Plant Biotechnology and Bioinformatics, U.Ghent, Dept. of Information Technology (INTEC, iMINDS), U.Ghent, Ghent, Belgium, (3)VIB Ghent University, Department of Plant Biotechnology and Bioinformatics, Ghent, Belgium

Eucalyptus has become an excellent model genus for studying lignocellulosic biomass accumulation due to the availability of a relatively small, high-confidence assembled genome, abundant gene expression studies and diverse species, populations and clonal collections. Through comparative genomics and transcriptomics approaches we have annotated most known genes and pathways known to be involved in xylogenesis. Despite these resources and years of xylogenesis research in *Arabidopsis thaliana* and *Populus* species and hybrids, we are still far from a complete understanding of the molecular regulation of carbon sequestration, allocation and partitioning for lignocellulosic biomass accumulation. To address this and to add comparative data for cross-species comparisons, we have produced multiple levels of quantitative biological data for carbon source, transport and sink tissues in three replicates from an *E. grandis* clone, including transcriptomics (mRNA and long non-coding RNA, small noncoding RNAs and miRNAs, proteomics and metabolomics. We also consider DNA- and important histone-modifications in these same samples. We integrate plastid and mitochondrial genomes and their encoded genes into this model, and the integration of these organelles' biology during wood formation, an important component rarely considered till now. Reconstruction of this and integration with accompanying systems genetics research in *Eucalyptus* populations is producing one of the most comprehensive models of wood formation and the regulation of lignocellulosic biomass accumulation.

W340: Forest Tree

The Chinese Chestnut Genome

Nathaniel Cannon, Penn State University, University Park, PA

Genome-wide marker-assisted selection can further the aims of the TACF back-cross breeding program by using Chinese chestnut (*Castanea mollissima*) as a reference to identify blight resistance genes introgressed into resistant American x Chinese hybrids. The Forest Health Initiative (http://www.foresthealthinitiative.org) provided support for the construction of a reference genome to meet this aim. Beginning in 2010 several years of sequencing obtained ~60x coverage of the genome yielding an assembly covering over 98% (784 Mbases) of the estimated size of Chinese chestnut genome. This draft assembly, released in 2014 (NCBI accession GCA_000763605.1) is represented by 42,000 scaffolds. These scaffolds were integrated with ~40,000 Sanger-end sequences from BAC clones which comprise a 1.8x physical map of the genome. BAC-ends are partially integrated with an additional ~1200 genetic markers which together served to order scaffolds into pseudochromosomes corresponding to 12 predicted linkage groups. Contigs from previously sequenced and assembled QTL regions have been incorporated into the pseudochromosome assembly. Additional sequencing has been under way since the release of the draft assembly in an effort to characterize sequence variation genome-wide, between and among Chinese and American chestnuts, and across the genus. Whole-genome sequence has been obtained for four additional Chinese genotypes, five American chestnut genotypes, and representative individuals from 6 additional species within the genus, all at approximately 7-10x depth. The draft assembly, pseudochromosomes, and QTL assemblies are available at the Hardwood Genomics website (http://hardwoodgenomics.org/chinese-chestnut-genome). Data tracks include supplementary data such as SNP calls and gene models, and are also available for download.

W341: Forest Tree

Functional Genomics of Poplar Bioenergy Phenotypes Using a Unique Dosage Variants Population

Héloïse Bastiaanse[1], Matthew S. Zinkgraf[1], Courtney Canning[1], Isabelle M. Henry[2], Luca Comai[3] and Andrew T. Groover[1], (1)US Forest Service, Davis, CA, (2)University of California, Davis, CA, (3)Plant Biology and Genome Center, UC Davis, Davis, CA

Poplars are among the fastest growing temperate trees in the world, with important ecological and economic value, including the production of biofuels and other value-added products. The poplar clones with the highest biomass yield are created through hybridization between species. Superior hybrid performance is believed to result in part from interactions among genomic regions that are in different copy number in the two species. Nevertheless the role of gene dosage in producing superior commercial hybrid poplar cultivars is not fully understood or effectively manipulated. We have produced a population of interspecific poplar hybrids carrying large-scale insertions and deletions (dosage lesions) scattered over the entire poplar genome. Dosage lesions were detected in approximately 55% of the progeny, and varied in length, position and number per individual, with a mean number of 2.4 lesions/individual and an average length of 5.7 Mb/lesion. Cumulatively, these lesions cover >99% of the genome with an average of 10 lesions affecting each gene. This population has been characterized for phenotypic traits related to biomass production, leaf morphology, wood properties and drought tolerance. Identification of genotypes displaying transgressive variation for specific traits will be described, as well as analyses seeking associations of phenotypic traits with genomic regions. Overall our study supports a better understanding of the role of gene dosage in poplar hybrid performance, and the identification of genes and genetic mechanisms contributing to poplar bioenergy traits. Additionally, since poplar trees can be clonally propagated, superior hybrid lines identified in this study are directly relevant to the industry.

W342: Forest Tree

Exploring Patterns of Sequence Variation in Regions Associated with Chestnut Blight Resistance Using Whole-Genome Resequencing of Chinese Chestnut (*Castanea mollissima*)

Nicholas LaBonte, Purdue University, West Lafayette, IN and Keith E. Woeste, USDA-FS HTIRC, Purdue University, west lafayette, IN

Chinese chestnut is the main donor of disease resistance in the blight resistance breeding program conducted by the American Chestnut Foundation (ACF), which aims to restore chestnut forests to eastern North America. Chinese chestnuts show variability in susceptibility to blight, but the genetic basis of this variability and its potential impact on the American chestnut breeding program have not been extensively studied. In particular, if differences in blight resistance correspond to high levels of allelic diversity at disease resistance loci in Chinese

chestnut, reliance on a small number of resistance donor parents would be problematic for the long-term success of the breeding program. The moderate size (~780-800 Mb) of the chestnut genome makes whole-genome resequencing a viable option for investigating functional genetic variation in Chinese chestnut. We have currently sequenced the whole genomes (Illumina 100 bp paired-end reads: depth 10-25x) of 16 Chinese chestnuts and interspecific hybrids with variable blight resistance, utilizing the Chinese chestnut draft genome sequence and the sequence of three major blight resistance QTL regions for assembly. We plan to sequence more individuals, and present preliminary analyses that have revealed patterns of sequence diversity across the three major blight resistance QTL and polymorphisms potentially associated with differences in blight resistance.

W343: Forest Tree

Modest Frequency and Pleiotropic Impacts of Zinc-Finger Mutagenesis in Poplar

Haiwei Lu, Amy Leigh Klocko, Michael Dow, Cathleen Ma, Vindhya Amarasinghe and Steven H. Strauss, Oregon State University, Corvallis, OR

Gene flow from recombinant-DNA modified (GMO) trees is major barrier to their public acceptance and regulatory approval. Because many intensively grown trees are vegetatively propagated, complete sexual sterility could be a powerful means to mitigate or prevent gene flow. We tested four pairs of zinc-finger nucleases (ZFNs) as mutagenic agents against the *LEAFY* and *AGAMOUS* orthologs in poplar that are expected to be required for sexual fertility. Each of the four *ZFN* genes was cloned behind a heat-shock promoter for inducible expression and reduced pleiotropic effects. Using *Agrobacterium tumefaciens*, we transformed more than 21,000 explants. The rate of transformation was substantially reduced by the ZFN constructs; only 391 transgenic shoots were produced (1.8 %). After heat shock and subsequent development of the transgenic plants, only two events were found to contain mutations; both were 7-bp deletions in one allele at the *PtAG2* locus. No mutations were observed at the *PtAG1* or *PtLFY* loci. Our results indicate a mutation rate of zero to 0.3% per explant per allele, among the lowest reported for ZFN mutagenesis in plants. The combined effects of reduced recovery of transgenic plants, and a modest mutation frequency, suggests that ZFNs may be a poor choice for mutagenesis of poplar genes.

W344: Fruit/Nuts

Functional Genomics as a Tool to Understand Cracking Tolerance in Sweet Cherry (*Prunus avium*)

Herman Silva[1], Juan Carlos Rios[1], Jonathan Maldonado[1], Lukas Schreiber[2], Viktoria Zeisler[2], Amit Dhingra[3] and Basilio Carrasco[4], (1)Lab Gen Func & Bioinfo - Fac Ciencias Agro - U de Chile, Santiago, Chile, (2)Institute of Cellular and Molecular Botany, Department of Ecophysiology, University of Bonn, Bonn, Germany, (3)Department of Horticulture, Washington State University, Pullman, WA, (4)Pontificia Universidad Católica de Chile, Santiago, Chile

The Chilean fruit industry has positioned itself as a key market for the country's development. To lead this market at the international level, it is necessary that Chile produce fruits and arrive to those markets with them in a very good condition/quality. Sweet cherries have become a fruit with a high value for the exporters and the growers. One of the problems associated with loss of production is cracking of sweet cherry fruit. Towards this end, it is essential to understand the molecular mechanisms that are involved in this disorder as well as identify genes and markers linked to this problem. In order to achieve this goal two approaches are under way. A biochemical approach was carried out to determine the role of alkenes in cracking using five different varieties of sweet cherry. After removal of cuticular wax, fruit cracking was significantly increased. Nuclear magnetic resonance analysis (^{1}H- and ^{13}C-NMR), revealed that fruits of different sweet cherry varieties contain primarily n-alkane with 29 carbons and no iso-alkane. Gas chromatography–mass spectrometry (GC-MS) enabled identification and quantification of n-alkanes. Varieties with significantly higher concentrations of nonacosane (Kordia, Regina and Lapins) were more tolerant to cracking compared to varieties with lower amounts (Bing and Rainier).

On the other hand, we used 454 Roche technology (454 GS-FL and 454 GS-FLX) to sequence the transcriptome of mature fruit from three sweet cherry varieties (Bing, Lapin and Rainier). Then, using CLC Genomics Workbench *de novo* assembly, we obtained a reference transcriptome of 20,349 contigs over 200 bp with a mean length of 919 bp. Illumina sequencing was performed for Bing variety under cracking water stress conditions. Several genes that might be involved in this disorder were identified. We have initiated a plum linkage map to validate and position the genes identified in sweet cherry.

This work is supported by CONICYT, FONDECYT/Regular Nº1120261

W345: Fruit/Nuts

Idiosyncratic Patterns of Disease Resistance Gene Adaptation in Three Species in the Rosaceae Detected Using the RosaR80 Framework

Leon Van Eck and James M. Bradeen, University of Minnesota, St. Paul, MN

Progress in identifying plant disease resistance genes is facilitated by the availability of high-quality reference genomes. We characterized the *R* gene composition of the apple (*Malus* × *domestica*), peach (*Prunus persica*) and woodland strawberry (*Fragaria vesca*) genomes using the RosaR80 framework, an analytical pipeline we developed as a community resource for comparative analysis in the Rosaceae. We detected species-specific patterns of NB-encoding gene adaptation: while woodland strawberry has wide representation of lineages from several ancient clades, peach and apple exhibit expansion of species-specific *R* gene lineages and large chromosomal gene clusters. We suggest an anthropogenic intensification of the molecular arms race as a driver of *R* gene lineage diversification. Compared to the wild woodland strawberry genome, the genomes of apple and peach represent varieties with prolonged cultivation as monocultures outside their native range. Domestication history influences *R* gene composition, and repeated selection for resistance to agriculturally relevant pathogens has driven expansion of these *R* gene lineages. We uncovered evidence for rapid adaptive diversification in response to pathogen pressure in the apple genome, where cloned and candidate genes affording resistance to apple scab (*Venturia inaequalis*) were derived from large, highly diversified, species-specific *R* gene lineages. Additionally, the RosaR80 framework revealed five lineages remarkably conserved between all three rosaceous species as well as phylogenetically distant plant species. Future research efforts will expand the RosaR80 framework, providing improved resolution for generating new evolutionary hypotheses, linking specific *R* gene lineages with specific putative functions, and informing evaluation of genebank collections for *R* allele discovery.

W346: Fruit/Nuts

DNA-Informed Strawberry Breeding in RosBREED

Sujeet Verma[1], Jack Roach[1], Jozer Mangandi[1], Seonghee Lee[1], Natalia Salinas[2], Nahla Bassil[3], Marco C.A.M. Bink[4], W. Eric Van de Weg[5], Cameron Peace[6], Amy Iezzoni[7] and Vance M. Whitaker[1], (1)University of Florida, Wimauma, FL, (2)Oregon State University, Corvallis, OR, (3)USDA/ARS, NCGR, Corvallis, OR, (4)Wageningen UR - Biometris, Wageningen, Netherlands, (5)Plant Breeding Wageningen University & Research centre, Wageningen, Netherlands, (6)Washington State University, Pullman, WA, (7)Michigan State University, East Lansing, MI

Providing breeders with DNA-based tools to combine disease resistance with excellent fruit quality in new rosaceous crop cultivars is the primary goal of the USDA-SCRI-funded RosBREED project. Here we present progress towards RosBREED objectives in the University of Florida (UF) strawberry breeding program. Large-effect QTLs for resistance to three strawberry diseases were identified: angular leaf spot caused by *Xanthomonas fragariae* (locus *Xf1* on LG6D), crown and root rot caused by *Phytophthora cactorum* (*Pc1* on LG7D), and crown rot caused by *Colletotrichum gloeosporioides* (*Cg1* on LG6B). In addition, moderate-effect QTL were identified for soluble solids content (LG6A) and fruit weight (LG6C and LG4B). These five QTLs were discovered in UF germplasm totaling more than 1,100 individuals from 122 full-sib families and validated in sets of advanced selections totaling more than 600 individuals. These individuals were phenotyped for disease resistance and fruit quality over two years and genotyped with the Affymetrix IStraw90 Axiom® SNP array. QTL discovery was conducted with FlexQTL™ software. Recombination patterns within QTL regions were analyzed and were used to inform SNP haplotype construction for QTL effect validations. Haplotype information was utilized in the selection of all parents and crosses in 2015. Marker-assisted seedling selection was performed on ~8000 seedlings using two previously developed DNA tests: a high-resolution melting assay for peach-like aroma (gamma-decalactone) and a SSR marker for perpetual flowering habit. Breeding gains realized with the use of DNA information and projected gains with the forthcoming DNA tests for *Xf1* and *Pc1* will be presented.

W347: Fruit/Nuts

Understanding Aphid Resistance in Black Raspberry, *Rubus occidentalis*

Jill M. Bushakra[1], Michael Dossett[2], Kelly Vining[3], Robert VanBuren[4], Doug Bryant[5], Jana Lee[6], Todd C. Mockler[5], Jungmin Lee[7], Chad Finn[8] and **Nahla Bassil**[1], (1)USDA/ARS, NCGR, Corvallis, OR, (2)Agriculture and Agri-Food Canada, Agassiz, BC, Canada, (3)Oregon State University, Corvallis, OR, (4)Danforth Plant Science Center, St Louis, MO, (5)Donald Danforth Plant Science Center, Saint Louis, MO, (6)USDA-ARS-HCRL, Corvallis, OR, (7)USDA-ARS, Parma, ID, (8)USDA-Agricultural Research Service, Corvallis, OR

The North American large raspberry aphid (*Amphorophora agathonica* Hottes) is the primary vector of Black raspberry necrosis virus (BRNV) which causes rapid decline in cultivated black raspberry (*Rubus occidentalis* L.) plant vigor leading to plant death and therefore poses a great risk to the viability of the Pacific Northwest black raspberry industry. The objective of this study was to understand the genetic architecture of three sources of aphid resistance identified in wild black raspberry germplasm collected from Ontario, Canada (ON), Maine (ME), and Michigan (MI). Linkage maps were constructed from three full-sib black raspberry populations designated ORUS 4304 (ME), ORUS 4305 (ON), and ORUS 4812 (MI) using microsatellite markers and single nucleotide polymorphisms generated with genotyping by sequencing. Preliminary mapping results indicate that all three loci are on the same linkage group, are near each other, and that each locus is unique. Phenotypic data from an additional 17 black raspberry populations with single and combinations of sources of aphid resistance have been collected. DNA-based markers are being developed for each of these sources to use for pyramiding multiple sources of resistance in new breeding material to ultimately provide resistance that will not easily be overcome by the aphids.

W348: Fruit/Nuts

Genome-wide annotation, tissue specific expression, and temporal expression profiling of transcription factors revealed candidate components of circadian clock in pineapple genome

Anupma Sharma[1], Ching Man Wai[2], Ray Ming[2] and Qingyi Yu[1], (1)Texas A&M AgriLife Research, Dallas, TX, (2)University of Illinois at Urbana-Champaign, Urbana, IL

Circadian clocks provide fitness advantage by coordinating internal metabolic processes and fine-tuning physiological events to external cyclic environments. Core clock components exhibit rhythmic changes in gene expression over a 24 hour period, and many of them are transcription factors (TFs) and transcription coregulators (TCs). In this study, we annotated TF/TCs in pineapple genome, and analyzed their tissue-specific and temporal expression pattern. We identified 1,398 TFs from 67 TF families and 80 TCs from 20 TC families. Nine TF families and two TC families were most abundant, accounting for approximately 50% of TF/TCs. Grouping of TF/TCs based on their combinatorial expression in leaf, flower, root, and fruit identified 581 TF/TCs upregulated in only one tissue, and 193 TF/TCs co-upregulated in more than one tissue. Majority of the genes in the latter group were co-upregulated in both leaf and flower, or in the root and fruit. Members 14 TF families and 1 TC family were significantly upregulated in only one tissue, and only one TF family was significantly upregulated in two tissues. Temporal expression profiles of pineapple TF/TCs in the green leaf tip and white leaf base tissues revealed diel peak in the expression of 547 TFs and 36 TCs. A total of 158 TF/TCs were cycling only in the white leaf base, 251 were cycling only in the green leaf tip, and 174 were cycling in both tissues. We identified a set of 66 TF/TCs whose expression was phase-locked between the green and white tissue, and this set was enriched in homologs of known Arabidopsis circadian system components. Our results suggest that phase-locked TF/TCs are prospective circadian system components, and their functional characterization may be instrumental in understanding the regulation of CAM photosynthesis in pineapple.

W349: Fruit/Nuts

Genome Database for Rosaceae: Updates and New Directions

Dorrie Main[1], Sook Jung[1], Chun-Huai Cheng[1], Taein Lee[1], Ping Zheng[1], Stephen P. Ficklin[1], Jodi L. Humann[1], Jing Yu[1], Anna Blenda[2], Kristin Scott[1], Michael Byrd[3], Lisa DeVetter[1], Katherine M. Evans[4], Cameron Peace[1], Ksenija Gasic[2], Mercy Olmstead[5],

Michael Coe[6] and Albert G. Abbott[7], (1)Washington State University, Pullman, WA, (2)Clemson University, Clemson, SC, (3)Erskine College, Due West, SC, (4)Washington State University, Wenatchee, WA, (5)University of Florida, Gainesville, FL, (6)Cedar Lake Research Group, Portland, OR, (7)Forest Health Research and Education Center, University of Kentucky, Lexington, KY

The Genome Database for Rosaceae (GDR, www.rosaceae.org) is the community database for basic, translational and applied research in almond, apple, apricot, blackberry, cherry, peach, pear, raspberry, rose and strawberry. Built using the Tripal platform, GDR provides an online portal of up to date, curated and integrated genomics, genetics and breeding data, combined with a suite of tools facilitating intuitive data mining and analysis. We highlight new data and functionality in GDR and indicate plans for further development through 2019 for this USDA and NSF funded resource.

W350: Fruit/Nuts

'HoneySweet', a New GE Plum Cultivar - From Concept to Product

Ralph Scorza[1], Ann Callahan[1], Christopher Dardick[1], Michael Ravelonandro[2], Mariano Cambra[3], Ioan Zagrai[4], Jaroslav Polak[5] and Tadeusz Malinowski[6], (1)USDA-ARS Appalachian Fruit Research Station, Kearneysville, WV, (2)INRA- UMR Fruit Biology and Pathology, Villenave d'Ornon, France, (3)IVIA, Valencia, Spain, (4)Fruit Research and Development Station Bistrita, Bistrita, Romania, (5)Crop Research Institute, Prague, Czech Republic, (6)Research Institute of Horticulture, Skierniewice, Poland

Genetic engineering (GE) has been intensively used to improve major agronomic crops but it has not been widely used in woody perennials, species that have the greatest need for this technology due to their protracted breeding cycles. The GE research and development (R&D) pipeline for woody perennials generally ends at proof-of-concept with few products entering the market. The USDA-ARS Appalachian Fruit Research Station fruit breeding program in collaboration with U.S. and European partners developed a GE approach to target resistance to *Plum Pox Virus* (PPV) the causal agent of Sharka, one of the most destructive diseases of stone fruits (*Prunus* sp.) This program developed 'HoneySweet', a PPV resistant, GE plum cultivar that has been tested for over 15 years in Europe and in the U.S. The 'HoneySweet' R&D pipeline included elucidation of transgene structure and expression, fruit compositional analyses, environmental risk assessments, tests of the durability of PPV resistance over time and under pressure of multiple *Prunus* viruses, effects on virus diversity and on non-target insects, and gene flow. These data moved 'HoneySweet' through the R&D pipeline to regulatory submissions. 'HoneySweet' was the first woody perennial to receive full regulatory approval in the U.S. 'HoneySweet' represents a new and novel source of PPV resistance available to growers and a germplasm resource for breeders with which to produce additional PPV resistant cultivars. The development, regulatory approval, and release of 'HoneySweet' plum provides an example of public institution GE R&D that offers practical solutions to challenges affecting agricultural production and food security.

W351: Fruit/Nuts

The Development of New Diploid (*F. iinumae*) and Polyploid Reference Genomes for Strawberry (*Fragaria*)

Lise L. Mahoney[1], Daniel J. Sargent[2], Yilong Yang[1], Dave J. Wood[1], Judson Ward[3], Nahla Bassil[4], James F. Hancock[5], Kevin M. Folta[6] and **Thomas M. Davis**[1], (1)University of New Hampshire, Durham, NH, (2)Center for Research and Innovation, Fondazione Edmund Mach, San Michele all' Adige, Trento, Italy, (3)Driscoll's, Watsonville, CA, (4)USDA/ARS, NCGR, Corvallis, OR, (5)Michigan State University, East Lansing, MI, (6)University of Florida, Gainesville, FL

The *F. vesca* 'Hawaii 4' reference genome has proven to be an immensely valuable resource for strawberry genetic/genomic research and resource development, including the IStraw90 SNP array. However, of the four pairs of subgenomes present in octoploid *Fragaria*, including the cultivated species *F. xananassa*, only one pair is derived from *F. vesca*. Here we describe the development of a second diploid strawberry reference genome, that of the ancestral diploid *F. iinumae* (Fii). Using germplasm collected in Hokkaido, Japan in 2004, we have developed an F2 generation mapping population which was subjected to SNP array and GBS genotyping, culminating in the construction of a high density linkage map of 4,173 markers at 496 loci spanning 451 cM. This map is now being used to anchor a Fii genome assembly. Comparison of the Fii linkage map to available versions of the 'Hawaii 4' genome assembly reveals no large scale genomic rearrangements, but illuminates discrepancies that may be indicative of small scale divergences and/or remaining errors in the latest 'Hawaii 4' assembly. The development of an octoploid reference genome is also being pursued, utilizing a novel strategy in which diploid x octoploid crosses are used to produce pentaploid progeny which can be exploited as "surrogate haploids". A preliminary linkage map has been constructed using SNP array data from a pentaploid progeny population, and whole genome sequencing of a specific pentaploid individual has been initiated.

W352: Fruit/Nuts

MicroRNA172 underlies apple fruit size evolution

Jia-Long Yao[1], Juan Xu[2], Amandine Cornille[3], Sumathi Tomes[1], Sakuntala Karunairetnam[1], Zhiwei Luo[1], Heather M Bassett[4], Claire Whitworth[5], Jonathan Rees-George[1], Chandra Ranatunga[5], Alodie Snirc[3], Ross N Crowhurst[1], Nihal De Silva[1], Satish Kumar[5], David Chagne[4], Vincent G.M. Bus[5], Richard K. Volz[5], Erik H. A. Rikkerink[1], Susan E Gardiner[4], Tatiana Giraud[3], Robin MacDiarmid[1] and Andrew P. Gleave[1], (1)The New Zealand Institute for Plant & Food Research Limited, Auckland, New Zealand, (2)Huazhong Agricultural University, Wuhan, China, (3)Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France, (4)The New Zealand Institute for Plant & Food Research Limited, Palmerston North, New Zealand, (5)The New Zealand Institute for Plant & Food Research Limited, Havelock North, New Zealand

The molecular genetic mechanisms underlying fruit size remain poorly understood in perennial crops, despite size being an important agronomic trait. Here we show that the expression level of a *microRNA* gene (*miRNA172)* influences fruit size in apple. A transposon insertional allele of *miRNA172* showing reduced expression associates with large fruit in an apple breeding population, whereas over-expression of *miRNA172* in transgenic apple significantly reduces fruit size. The transposon insertional allele was found to be co-located with a major fruit size quantitative trait locus, fixed in cultivated apples and their wild progenitor species with relatively large fruit. This finding supports the view that the selection for large size in apple fruit was initiated prior to apple domestication, likely by large

mammals, before being subsequently strengthened by humans, and also helps to explain why signatures of genetic bottlenecks and selective sweeps are normally weaker in perennial crops than in annual crops.

W353: Fruit/Nuts

A Draft Assembly of the Almond Genome

Tyler S. Alioto[1], Fernando Cruz[1], Jèssica Gómez-Garrido[1], Leonor Frias[1], Paolo Ribeca[1], Marta Gut[1], Konstantinos Alexiou[2], Werner Howad[2], Jordi Morata[3], Josep Casacuberta[3], Maria J. Rubio-Cabetas[4], Beatriz Galán[5], José Luís García[5], Amit Dhingra[6], Henri Duval[7], Angel Fernandez i Marti[8], Michelle Wirthensohn[9], Ivo G. Gut[1] and Pere Arús[2], (1)Centro Nacional de Análisis Genómico (CNAG-CRG), Centre for Genomic Regulation, Barcelona, Spain, (2)IRTA, Center for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Bellaterra, Spain, (3)Center for Research in Agricultural Genomics CRAG (CSIC-IRTA-UAB-UB), Bellaterra, Spain, (4)Hortofruticulture Department, Agrifood Research Center of Aragón (CITA), Zaragoza, Spain, (5)Environmental Biology Department, Centro de Investigaciones Biológicas, (CSIC), Madrid, Spain, (6)Department of Horticulture, Washington State University, Pullman, WA, (7)INRA, UR1052, Unité de Génétique et Amélioration des Fruits et Légumes (GAFL), Montfavet cedex, France, (8)Genome Center, University of California, Davis, CA, (9)University of Adelaide, School of Agriculture, Food and Wine, Glen Osmond, SA, Australia

Almond is one of the oldest cultivated nut crops with its origin in central and western Asia. The selection of the sweet type (*Prunus dulcis*) distinguishes the domesticated almond from its bitter wild relatives. It is economically important, especially in California with the highest worldwide production, followed by Australia and Spain. The almond belongs to the same subgenus as the peach, for which there already exists a reference genome. However, to fully understand the genetic underpinnings marking the key phenotypic differences between almond and peach, we have sequenced the genome of the 'Texas' almond, one of the traditional cultivars producing a sweet nut. Whole-genome shotgun sequencing of Illumina paired-end libraries gave an initial low-contiguity assembly of 512 Mbp, nearly double the estimated genome size. Counting of k-mers indicates a 275 Mbp genome with substantial heterozygosity as well as repetitive sequence. In order to tackle both problems, we constructed a fosmid library and sequenced 68 pools of ~500 clones per pool. We then assembled the pools, merged them and finished the assembly by scaffolding with paired end and mate pair libraries, which resulted in a 240 Mbp assembly with a scaffold N50 of 500 kbp, a contig N50 of 33.5 kbp and CEGMA completeness of 99%. Two thirds of the assembly was anchored to the peach-almond genetic map, and using re-sequencing data of peach-almond hybrids and their parents we inferred the two haplotypes of the sequenced almond tree. We performed additional validation of the assembly using Oxford Nanopore MinION sequencing.

W354: Fruit/Nuts

RNA-Seq Analysis and Annotation of a Draft Blueberry Genome Assembly

Ann Loraine[1], Rob Reid[1] and Allan Brown[2], (1)University of North Carolina Charlotte, Kannapolis, NC, (2)International Institute of Tropical Agriculture, Arusha, Tanzania, United Republic of

Highbush blueberry is an important small fruit crop in the US, Chile, New Zealand, and many other countries. Thanks to decades of breeding, blueberries can now be grown in many different locations and climates. Understanding the genetic and molecular basis for valuable berry traits requires access to genome sequence and annotations. We used a combination of sequencing strategies to generate a draft genome assembly for blueberry. We annotated the assembly using RNA-Seq data from five stages of berry fruit development and ripening. Homology-based annotation assigned potential functions to more than 60% of gene models. RNA-Seq expression profiling showed that blueberry growth, maturation, and ripening involve dynamic gene expression changes, including coordinated up- and down-regulation of metabolic pathway enzymes and transcriptional regulators. Analysis of RNA-seq alignments identified developmentally regulated alternative splicing, promoter use, and 3′ end formation. Data are available for visualization in Integrated Genome Browser via an IGB Quickload site located at http://www.igbquickload.org/blueberry/. In this talk, we'll describe these resources and highlight major results obtained from genome annotation and RNA-Seq analysis of berry fruit.

W355: Fruit/Nuts

Wild and Cultivated Apricot Genetic Structure Linked to Plum Pox Potyvirus Resistance

Veronique Decroocq[1], Stéphane Decroocq[2], Amandine Cornille[3], David Tricon[2], Shuo Liu[4], Wei-Sheng Liu[5], Tatiana Giraud[6] and Albert G. Abbott[7], (1)UMR BFP1332 - INRA-Universite de Bordeaux, Villenave d'Ornon Cedex, France, (2)UMR BFP1332 - INRA-Universite de Bordeaux, Villenave d'Ornon, France, (3)Swiss Federal Institute of Technology of Zürich ETHZ, Zürich, Switzerland, (4)Institut National de la Recherche Agronomique, Université de Bordeaux, UMR BFP 1332, Villenave d'Ornon, France, (5)Liaoning Institute of Pomology, Xiongyue, Yingkou City Liaoning, China, (6)Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France, (7)Forest Health Research and Education Center, University of Kentucky, Lexington, KY

Apricot (*Prunus armeniaca* L.) is an important horticultural stone fruit species in the Northern hemisphere, where it is being severely threatened by *Plum Pox Virus* (PPV) infection (sharka disease). In a previous genome-wide association study (Mariette et al., New Phytologist, 2015), we identified PPV resistant apricot cultivars and mapped the resistance determinants. This work was completed, in the frame of the FP7 Marie Curie STONE project (#246795), by a search of the geographical origin and a world-wide genetic diversity analysis of resistance source(s) to sharka in apricot. Indeed, the apricot wild progenitor, *P. armeniaca vulgaris*, is still present as a forest tree in its native area, on the slopes of the Tien Shan range, where it co-occurs with the wild apple species, *Malus sieversii*. We thus investigated patterns of molecular diversity and genetic structure using microsatellite markers in a dataset of 380 wild apricot accessions collected in Central Asia, natural populations (Kazakhstan, Uzbekistan, Western China and Kyrgyzstan). They were compared with Western and Eastern cultivated relatives in order to determine putative geographic origin of European cultivated apricots and to reconstruct its evolutionary history. We observed high levels of genetic diversity in Central Asian and Chinese germplasm, in agreement with an origin of this species in this region. Lastly, we unraveled the origin of resistance to PPV and present here preliminary results on its phenotypic variability within the native apricot gene-pool, in comparison with European and North-American cultivated apricots.

W356: Fruit/Nuts

Mapping Human Taste Perception on the Apple Genome

Daryl J. Somers[1], Beatrice Amyotte[1], Amy Bowen[1], Travis W. Banks[1], David Liscombe[1] and Istvan Rajcan[2], (1)Vineland Research and Innovation Centre, Vineland Station, ON, Canada, (2)University of Guelph, Guelph, ON, Canada

Breeding apples can take 10-20 years before elite cultivars are selected and tested for production and commercial value. In 2011, Vineland initiated an apple breeding program in partnership with the Ontario Apple Growers. Consultation with the apple sector strongly emphasized the importance of breeding high quality, marketable apples. Vineland launched an apple project to merge human sensory perception, consumer preference, analytical evaluation and apple genome sequencing with a long term goal of mapping the human taste experience onto the apple genome. This would give the breeding program valuable insight into genomic regions and DNA markers that are associated with high quality attributes of apple that are preferred by consumers. A collection of 80 apple accessions were examined over 2 years by a human sensory panel measuring 18 attributes, a consumer taste panel to measure preference, 5 analytical measures and Genotype-by-Sequencing (GBS). The genome wide association analysis showed human sensory traits could be mapped including juciness, crispness, mealiness, and fresh apple flavour with QTL explaining between 20-35% of the trait variation. The research also leads to a search for causative genes and SNPs underlying the mapped traits. This presentation will summarize the current findings from the project and discuss next steps plus applications of this approach in other crop species.

W357: Functional Annotations of Animal Genomes (FAANG)

Brief Introduction of FAANG

Huaijun Zhou, University of California, Davis, Davis, CA and Christopher K. Tuggle, Iowa State University, Ames, IA

W358: Functional Annotations of Animal Genomes (FAANG)

Brief Introduction of RCN

Fiona McCarthy, College of Agriculture and Life Sciences, Tucson, AZ

W359: Functional Annotations of Animal Genomes (FAANG)

Metadata and Data Sharing Committee report

Laura Clarke, European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, United Kingdom and Carl J. Schmidt, University of Delaware, Newark, DE

The Functional Annotation of ANimal Genomes (FAANG) initative aims to create a comprehensive functional annotation for many animal genomes. A key requirement is to define strong metadata standards ensuring the data is well described and making the raw data maximally useful. It is also important that data can be rapidly shared both within FAANG and with the wider community.

The Metadata and Data Sharing Committee has three main goals:

1. Define metadata standards for samples, experiments and analyses.
2. Provide guidance for archival best practice.
3. Enable data sharing within FAANG and with the wider community.

Our metadata standards cover: samples, with information for both donor animals and biological material; experiments, with information on tissue/cell type derivation, sequence extraction and sequencing protocol; and analyses detailing alignment, quantification and signal processing protocols. These standards are versioned in a github repository.

All members of FAANG commit to following the Toronto Data Release Workshop guidelines[1]; the Metadata and Data Sharing group are writing a data archiving strategy to help with this aim. This will provide guidance about the best assay archives to store different data types in, ensuring FAANG data is well represented in the assay archives hosted at EMBL-EBI and NCBI.

Finally the Metadata and Data sharing group will provide both within consortium data exchange and a public FTP site to ensure the data is rapidly and easily available to the whole community.

1. Toronto International Data Release Workshop Authors, Birney E, et al. Prepublication data sharing. Nature. 2009: 461:168-70.

W360: Functional Annotations of Animal Genomes (FAANG)

Animals, Samples, and Assay Committee report

Elisabetta Giuffra, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France and Huaijun Zhou, University of California, Davis, Davis, CA

Standardization of every assay/protocol involved in the functional annotation of animal genomes is an essential prerequisite for sharing, integrating and global analysis of multiple datasets. The overall objective of the Animals/Samples/Assays (ASA) Committee is to propose, develop and standardize animal/tissue collection protocols, storage practices, and processing protocols for the FAANG community.

Concerning animal/sample collections, breed, age, environment and physiological state at sampling vary, which increases heterogeneity but also the potential for comparisons across tissues and cell types. To organize FAANG collections and deal with this complexity, ASA collaborates with the 'Metadata & Data Sharing' Committee for defining the minimum requirements, suitable ontologies, reference databases and example metadata for the FAANG collections.

Protocols for molecular assays are being developed based on standards defined by the ENCODE projects and the International Human Epigenome Consortium (IHEC) as references. A set of core assays, which primarily employ technologies (whole genome re-sequencing, RNA sequencing, chromatin accessibility and histone marks) that work across all targeted species are the starting point; additional assays are being developed by individual research groups based on specific needs and research interests (Box 1 in Andersson et al. 2015; PMID: 25854118). Committee's members are working to adapt these assays where necessary, taking into account the complexities of individual species and the different tissues available. Once drafted, protocols are shared through the Committee's page of the FAANG wiki

(http://www.ebi.ac.uk/seqdb/confluence/display/FAANG/) and finally published on the ASA Committee webpage as proposed standards to the FAANG community.
The expected outcome of these activities is to standardize the best practice tools and resources for facing the genotype-to-phenotype challenge in domesticated animals.

W361: Functional Annotations of Animal Genomes (FAANG)
Bioinformatics and Data Analysis Committee report
James M. Reecy, Iowa State University, Ames, IA and Mick Watson, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom

W362: Functional Annotations of Animal Genomes (FAANG)
FAANG Associated Projects Update
Martien A.M. Groenen, Wageningen University, Wageningen, Netherlands and Graham S. Plastow, University of Alberta, Edmonton, AB, Canada

W363: Functional Annotations of Animal Genomes (FAANG)
Follow up discussion
Graham S. Plastow, University of Alberta, Edmonton, AB, Canada and Elisabetta Giuffra, INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France

W364: Functional Genomics
High-throughput field phenotyping: bridging scales from gene to canopy for trait discovery
David Gouache[1], Benoit de Solan[2], Antoine Fournier[3], Alexis Comar[4], Fred Baret[5], Stéphane Lafarge[6], Agathe Mini[7], Benoit Piquemal[8], Matthieu Bogard[9], Xavier Lacaze[9], Rongling Wu[10] and Katia Beauchene[3], (1)ARVALIS, Boigneville, France, (2)ARVALIS - Institut du végétal, Avignon, France, (3)ARVALIS - Institut du végétal, Ouzouer le marché, France, (4)HiPhen, Avignon, France, (5)INRA, Avignon, France, (6)Biogemma, Chappes, France, (7)ARVALIS - Institut du végétal, Chappes, France, (8)ARVALIS - Institut du végétal, Boigneville, France, (9)ARVALIS - Institut du végétal, Baziège, France, (10)Pennsylvania State University, Hershey, PA
Recent improvements in phenomics have opened up vast opportunities to understand and predict dynamic phenotypic responses to genotypic variation. To date however, a great majority of phenomics studies are conducted in controlled condition facilities, such as growth chambers and greenhouses, in which plants function as individuals. In crop species however, plants function as canopies. This drastically changes the way they function and limits the relevance of some controlled condition studies. Thankfully, field phenomics platforms are emerging. We will present the results of a prototype developed to study the response of wheat genotypes to nitrogen in France. This system, based on RGB cameras and reflectance spectroradiometers, characterized the dynamic canopy response of over 200 wheat genotypes throughout the growing season. Analysis of the phenotypic dataset generated shows that traits relative to canopy structure and function are linked to the overall yield response of wheat to nitrogen. Preliminary genome wide association and functional mapping of these traits is underway and shows promise in identifying how genetic variation modifies canopy traits. These traits should contribute to linking results at lower scales to economically important crop variables such as yield. The accessibility of the phenotyping methods presented is increasing, thanks to the development of higher-throughput and more automated systems.

W365: Functional Genomics
Comparative Transcriptomics of Parasitic Plants Reveals Core Parasitism Genes and Elucidate Origins of Haustoria
Zhenzhen Yang, Penn State University, University Park, PA
The origin of novel traits is recognized as an important process underlying many major evolutionary radiations. We studied the genetic basis for the evolution of haustoria, the novel feeding organs of parasitic flowering plants, using comparative transcriptome sequencing in three species of Orobanchaceae. Around 180 genes are upregulated during haustorial development following host attachment in at least two species, and these are enriched in proteases, cell wall modifying enzymes, and extracellular secretion proteins. Additionally, about 100 shared genes are upregulated in response to haustorium inducing factors prior to host attachment. Collectively, we refer to these newly identified genes as putative "*parasitism genes*." Most of these parasitism genes are derived from gene duplications in a common ancestor of Orobanchaceae and *Mimulus guttatus*, a related nonparasitic plant. Additionally, the signature of relaxed purifying selection and/or adaptive evolution at specific sites was detected in many haustorial genes, and may play an important role in parasite evolution. Comparative analysis of gene expression patterns in parasitic and nonparasitic angiosperms suggests that parasitism genes are derived primarily from root and floral tissues, but with some genes co-opted from other tissues. Gene duplication, often taking place in a nonparasitic ancestor of Orobanchaceae, followed by regulatory neofunctionalization, was an important process in the origin of parasitic haustoria.

W366: Functional Genomics
Epigenetic Regulators of Gene Expression in Bean-Bean Rust Interaction
Venkateswara R. Sripathi, Center for Molecular Biology, Alabama A&M University, Normal, AL
Common bean (*Phaseolus vulgaris* L.) is economically important for its high protein, fiber and micronutrient contents. Bean-rust, caused by the fungal pathogen *Uromyces appendiculatus* is a major constraint in common bean production. Integrating the information extracted from next generation sequencing (NGS) approaches aid in understanding regulatory mechanisms associated with disease-resistance in common bean. This study utilized chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA sequencing (RNA-Seq) of inoculated (I) and mock-inoculated (MI) leaf samples that were collected at 0, 12 and 84 hour-after-inoculation (hai). The bioinformatic tools used for these analyses were Bowtie,

SICER, HOMER, TopHat and Cufflinks. We identified five important classes of epigenetic modulators in gene expression that include DNA methylation (279), histone methylation (45), histone acetylation (145), chromatin remodeling (26) and Polycomb Group (225) proteins. Four of the differentially expressed stress responsive genes identified were early-responsive to dehydration (ERD) stress family protein, chloroplast drought-induced stress protein, oxidative stress 3 and stress induced alpha-beta barrel domain protein between 0, 12 and 84 hai. The R proteins marked by both methylation and acetylation modifications were pleiotropic drug resistant protein 12, LRR family, NB-ARC domain containing and TIR-NBS-LRR proteins. Some key defense responsive genes (calmodulin, cytochrome p450, chitinase, DNA Pol II, and LRR) and transcription factors (WRKY, bZIP, MYB, HSFB3, GRAS, NAC, and NMRA) were also identified. This is the first comprehensive report in understanding the role of epigenetic regulation in bean-rust interaction and can be exploited in other non-model species under biotic and abiotic stresses.

W367: Functional Genomics

Soybean Mutations Mapping: Applications in Functional gene analysis and Soybean Improvement

Naoufal Lakhssassi, Shiming Liu, Zhou Zhou and Khalid Meksem, Department of Plant Soil and Agricultural Systems, SIUC, Carbondale, IL

Soybean [*Glycine max* (L.) Merr.] is the most widely consumed legume crop in the world, providing about 68% of world protein supply as well as food oil and renewable fuels. We developed several EMS mutagenesis soybean populations (Cooper et al., 2008; Meksem et al., 2008; Liu et al., 2012), and did a forward screening of these populations to identify genes of economical importance. We identified several lines with altered phenotype in specific traits, mainly oil, protein, high nodulating lines, high yielding and soybean cyst nematode resistance. A subset of the mutagenized population containing about 1000 families (M3 and M4 lines), was screened using reverse genetic approach: TILLING, the following mutants were identified within the omega-6 fatty acid desaturase gene *FAD2*, one silent mutation: F1274 (L249=), two missense mutation: F1284 (P284L) and F1796 (Q22R) in the *FAD2-1A* isoform. Another missense mutation: F812 (P163S) was identified in the other *FAD2-1B* isoform, all the four mutations resulted in an increase in oleic acid content up to 2.5 times compared to the wild type parent Forrest.

Using a forward screening approach, we were able to identify four mutants: F605, F620, F714, and F813, all of which contain high level of seed stearic acid with an increased level of up to 2.4 times compared to the wild type parent Forrest used to develop the mutagenized population. Each of these four mutants was identified to carry one missense or nonsense mutation of *SACPD-C*: Q83* (F605), L79F (F620), D77N (F714), and P102L (F813), among them, Q83*, L79F, and P102L are the new *SACPD-C* alleles identified in soybean.

The identified mutants can be used as new SCN resistant soybean sources with high seed stearic and oleic acid contents for further breeding purposes.

Keywords: Soybean, EMS mutagenesis, mutants population, mutation breeding, fatty acids, stearic acid, *SACP-D*, *SACPD-C*, *SACP-B*, soybean cyst nematode.

W368: Functional Genomics

Identification of QTL Controlling Symbiotic Nitrogen Fixation in Soybean

Francois Belzile, Laval University, Québec, QC, Canada

Symbiotic nitrogen fixation (SNF) allows legumes to benefit from the conversion of atmospheric nitrogen into forms of nitrogen that are useable by the plant thanks to a complex symbiotic relationship with rhizobacteria. Despite the importance of SNF in soybean, very few studies have been conducted to dissect the genetic architecture of SNF in this crop. We genotyped 292 lines representative of the International Institute for Tropical Agriculture (IITA) soybean germplasm with 50K high-quality SNPs obtained via genotyping-by-sequencing (GBS). Greenhouse-grown plants were either inoculated with a single strain of *Bradyrhizobium japonicum* or fed with nitrogen fertilizer. A total of nine traits related to nitrogen fixation were measured: Shoot Dry Weight (SDW), Root Dry Weight (RDW), Number of Nodules (NN), Nodule Dry Weight (NDW), NDW/SDW ratio, normalized NDW (nNDW), NDW/RDW ratio, $SDW_{inoculated}/SDW_{fertilized}$, and $RDW_{inoculated}/RDW_{fertilized}$. In total, GWAS analysis identified over 50 highly significant QTL regions ($q < 0.01$). Among these, a highly significant association between markers on Gm19 and SDW (p=6.45E-06) was noted, but this association was found in both inoculated and fertilized plants. We observed that the RDW ratio between inoculated and fertilized plants ($RDW_{inoculated}/RDW_{fertilized}$) proved the most informative in terms of detecting QTLs unique to the inoculated treatment. In the interval extending 100 kb to either side of the peak marker of our QTLs, we identified 15 genes previously reported to be involved in SNF. Compared to the genome-wide occurrence of such genes, this represents an 80-fold enrichment (p=2.93E-16). We hope our results will shed further light on SNF in soybean.

W369: Functional Genomics

RESYNTHESIS AND VORIs: New Marker-Based Approaches for Peach Breeding

Pere Arús, IRTA (Institut de Recerca i Tecnologia Agroalimentàries), Bellaterra, Spain

W370: Functional Genomics

The Azolla Metagenome

Laura Dijkhuizen[1], Paul Brouwer[1], Henk Bolhuis[2], Gert-Jan Reichert[1], Nils Koppers[3], Bruno Huettel[4], Fay-Wei Li[5], Xin Liu[6], Gane Ka-Shu Wong[7], Kathleen M. Pryer[8], Andreas Weber[9], Andrea Bräutigam[9] and **Henriette Schluepmann**[1], (1)Utrecht University, Utrecht, Netherlands, (2)Netherlands Institute for Sea Research (NIOZ), Yerseke, Netherlands, (3)Heinrich Heine University, Düsseldorf, Germany, (4)Max Planck Genome Centre, Cologne, Germany, (5)University of California Berkeley, Berkeley, CA, (6)Beijing Genomics Institute-Shenzhen, Shenzhen, China, (7)Department of Biological Sciences, University of Alberta, Edmonton, AB, AB, Canada, (8)Duke University, Durham, NC, (9)Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

Ferns from the genus *Azolla* are floating invasive weeds in waterways worldwide, yet industrialized biomass production with *Azolla* may be advantageous not the least because of their N_2-fixing symbionts, *Nostoc azollae*. Experimental production with *A. filiculoides* in the absence of N-fertilizer yielded over 1200 kg N $ha^{-1}yr^{-1}$ fixed in the biomass harvested implying that *Azolla* should be considered as forage to substitute

soybean imports in European latitudes where water is available. The genome of *A. filiculoides* was sequenced, assembled and annotated towards domestication and breeding, yet we wondered about biosafety of the biomass and therefore began to characterize the metagenome of *A. filiculoides* from a Dutch ditch and the corresponding sterilized strain grown in the laboratory. The approach included metagenome sequencing, taxonomic assignments using rRNA gene sequences, assembly of genome sequences and recruitment analyses. In addition, nitrogen cycle enzymes in the assembled persistent *Rhizobium* endophyte were annotated, and $^{15}N_2$ fixation and biomass $\delta^{15}N$ measured. We conclude 1) the immune system of *A. filiculoides* is very selective, 2) *Azolla*'s main symbiont in the leaf pockets is *N. azollae*, 3) denitrifying *Rhizobia* are generally associated with *Azolla* species in culture or in *A. filiculoides* from the ditch 4) *Rhizobia* are enriched in leaf pockets and may detoxify excess NH_4^+ produced by *N. azollae* or present in waterlogged anoxic sediments where nitrification is inhibited and NH_4^+becomes the primary N-source

W371: Functional Genomics

Lessons From Whole Genome Sequencing of Olive Tree (Olea europaea L.)

Turgay Unver, Biology Department, Faculty of Science, Cankiri Karatekin University, Cankiri, Turkey, Mine Turktas, Cankiri Karatekin University, Cankiri, Turkey, Gabriel Dorado, Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, Cordoba, Spain, Pilar Hernandez, Instituto de Agricultura Sostenible, Cordoba, Spain, Yves Van de Peer, VIB - Ghent University, Zwijnaarde, Belgium and International Olive Genome Sequencing Consortium, IOGC, Cankiri, Turkey

The olive tree (*Olea europaea* L.) is an economically-important fruit tree, being widely spread in the Mediterranean Basin. It is one of the most-important oil crops in the world, belonging to the family of *Oleaceae*, order of Lamiales, which includes important members for their essential oils, fragrances and phenolics. Olive tree is a diploid (2n = 2x = 46), predominantly allogamous, and vegetatively propagated species. To date, there is no available genome-level sequence data of olive tree. The wild olive-tree (*O. europaea* var. *sylvestris*) genome was sequenced and assembled with 246X coverage. SOAPdenovo was used to assemble reads, which resulted in a draft genome of 1.48 Gbp with scaffold N50 of 228 kbp, which is near to previous estimates by flow cytometry and *k*-mer analyses (~1.46 Gbp). A total of 42,843 scaffolds (>1 kbp) were assembled with about 80% of the total assembly (1.14 Gbp). Using a newly-constructed genetic map, 50% of the sequences were anchored into 23 linkage groups, which included 572 Mbp. About 50% of the total genome assembly was found to be composed of repetitive DNA. Transposable elements and interspersed repeats occupied 50% of the genome. Protein-encoding gene models were constructed by combined methodology, including *de novo*, homology-based and RNA-Seq-aided predictions. A total of 50,684 protein-encoding gene models were predicted for total assembly, of which 31,245 were anchored into chromosomes. To estimate heterozygosity in the genome, a *k-mer* distribution approach was performed, showing a distribution of about 1.3%. A phylogenetic tree was constructed for evolutionary analyses, on the basis of a concatenated sequence-alignment of the 99 single copy genes shared by olive tree and other 12 plant species. In this phylogenetic dendrogram, the olive-tree clustered with oil-crop sesame. Whole-genome duplication and speciation events were analyzed, via fourfold synonymous third-codon transversion (4DTv) approach, showing that there was a recent whole-genome duplication event in the olive-tree, before speciation from sesame. The olive tree genome was also subjected to synteny analyses with sesame, resulting in 1,727 synteny blocks. Genes involved in oil biosynthesis, ripening, secondary metabolite synthesis and alternate bearing, as well as small RNAs and transcription factor genes were annotated and further analyzed, using transcriptome data from several tissues. The olive-tree reference genome will serve as a crucial source not only for the study of olive-tree genome biology, but also for tree genomics, further facilitating more effective olive tree breeding programs.

W372: Functional Genomics

Whole genome functional annotation of wild olive-tree (Olea europaea var. sylvestris)

Oussama Badad[1], David A. Lightfoot[2], Hassan Ghazal[3], Gabriel Dorado[4], Pilar Hernandez[5], Mine Turktas[6], Turgay Unver[6] and International Olive Genome Sequencing Consortium[7], (1)University Mohamed The Fifth, Rabat, Morocco, (2)Southern Illinois University, Carbondale, IL, (3)University Mohamed Premier, oujda/Nador, Morocco, (4)Campus de Excelencia Internacional Agroalimentario (ceiA3), Universidad de Córdoba, Cordoba, Spain, (5)Instituto de Agricultura Sostenible, Cordoba, Spain, (6)Cankiri Karatekin University, Cankiri, Turkey, (7)International Olive Genome Sequencing Consortium, Cankiri, KS, Turkey

Olive (Olea europaea L.) is an economically-important fruit tree, widely spread in the Mediterranean Basin, and belonging to the family of Oleaceae, order of Lamiales, which includes important members for their essential oils, fragrances and phenolics. The olive tree is a diploid (2n = 2x = 46), predominantly allogamous, and vegetatively-propagated species. The genome of the wild olive-tree (O. europaea var. sylvestris) has been recently sequenced and assembled. Structural annotation revealed 60,214 protein-encoding genes (CDS). For the functional annotation, we adopted a double approach: a BLAST-based approach and a protein-domain search approach (Blast_v2.2.30p.sh and interproscan5.sh). The results were combined on a Blas2Go pro project for further Gene Ontology (GO) analysis and functional description assignments. From the total of 60,214 CDS sequences, 23% fail to obtain a BLAST hit, 4% of the BLASTed sequences without hits with functional information cannot be linked to GO entries and 9% of the sequences with GO mapping does not reach the quality for an annotation assignment. The high percentage of CDS with no BLAST hits would be due to the quality of assembly or structural annotation. Overall, we could assign functional labels to 64% of the input sequences. In addition, the Enzyme Code (EC) and KEGG analysis allowed the identification of 1,480 EC out of 133 KEGG pathways. These results are similar to the ones obtained from other sequenced plant genomes, using the same approach. Meanwhile, we are developing the olive genome browser, as platform for data visualization (chromosome map, gene structure, repeats, sRNA annotations) .

W373: Functional Genomics of C_4 and CAM photosynthesis

Rapid Optical Profilometry and Computer Vision of Leaf Epidermal Structure Applied to Genetic and Environmental Control of Stomatal Patterning in Model C_4 Species

Andrew D.B Leakey[1], Gorka Erice[2], Nathan Miller[3], Dan Xie[4], Max J Feldman[5], Matthew Kendzior[4], Parthiban Prakash[4], Elizabeth A. Ainsworth[6], Ivan Baxter[7], Patrick J. Brown[1], Michael V. Mickelbart[8] and Edgar Spalding[9], (1)University of Illinois, Urbana-Champaign, Urbana, IL, (2)University of Illinois at Urbana-Champaign, Urbana, IL, (3)University of Wisconsin, Madison, WI, (4)University of Illinois, Urbana, IL, (5)Donald Danforth Center Plant Science Center, St. Louis, MO, (6)USDA-ARS, Urbana, IL, (7)USDA-ARS, Danforth Plant Science Center, St. Louis, MO, (8)Department of Botany and Plant Pathology, Department of

Horticulture and Landscape Architecture, Purdue University, West Lafayette, IN, (9)University of Wisconsin Madison, Madison, WI

Leaf epidermal structures, including stomata and hairs, play key roles in leaf function. Stomatal and hair patterning are highly regulated developmental processes in response to both environmental and genetic signals. Modern quantitative genetics approaches have not been fully applied to understanding epidermal structures due the laborious nature of phenotyping methods. C_4 grasses are agriculturally and ecologically important, in large part due to their high water use efficiency. Yet, little is known about the mechanisms controlling stomatal and hair patterning in this key plant functional type. We have developed and applied a rapid method of assessing stomatal and hair patterning in two model C_4 species – maize and setaria. The leaf surface is scanned in less than two minutes with a modified confocal microscope, generating a quantitative measurement of a patch of the leaf surface. We have developed an algorithm for automatically detecting stomata in epidermal surfaces through training of a pattern-recognition neural network. We have validated this rapid phenotyping technique in: (1) diverse *Zea mays* inbreds grown at ambient and elevated ozone using free-air concentration enrichment (FACE) technology; and (2) in a recombinant inbred line (RIL) population resulting from the cross of *Setaria italica* x *Setaria viridis*. Variation in stomatal patterning among founder lines of the NAM population of *Z. mays* was reproducible between field and greenhouse conditions at Illinois and Purdue. SD was reduced in a subset of maize genotypes grown at elevated ozone compared to ambient ozone. QTL for stomatal patterning and surface roughness were identified in Setaria.

W374: Functional Genomics of C_4 and CAM photosynthesis

Tissue Succulence Engineering for CAM Biodesign

Sung Don Lim, Won Cheol Yim and John C. Cushman, Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV

Crassulacean acid metabolism (CAM) is a photosynthetic adaptation present in 6-7% of vascular plant species that reduces water loss through stomata by shifting the diel rhythm of primary atmospheric CO_2 fixation from the day to the night, which results in improved water-use efficiency (WUE) and drought tolerance. Introducing the CAM photosynthetic machinery into C_3 plants (CAM biodesign) is expected to confer improved WUE in order to allow plants to withstand long episodes of drought or perhaps even assist in the expansion of crop production into semi-arid regions. Some degree of tissue succulence is typically correlated with the optimal performance of CAM. Thus, for CAM biodesign, tissue succulence engineering in C_3 plants represents a key anatomical prerequisite for enhancing the efficient operation of engineered CAM in C_3 photosynthesis species. Tissue succulence engineering is expected to afford increased mesophyll cell size in order to increase malate storage capacity in the vacuoles of the larger cells and to reduce intercellular air space (IAS) to limit the diffusion of CO_2 out of the leaf upon its release during day. A method for the genetic engineering of tissue succulence was developed involving the overexpression of a modified basic helix-loop-helix (bHLHL) transcription factor from *Vitis vinifera*. The engineered tissue succulence in *Arabidopsis* results in reduced IAS and increased cell size, organ size, organ number, vegetative biomass, and seed production.

W375: Functional Genomics of C_4 and CAM photosynthesis

Adaptive Evolution of C4 Photosynthesis

Pu Huang[1], Tony Studer[2], James C Schnable[3], Elizabeth A. Kellogg[1] and Thomas P. Brutnell[4], (1)Donald Danforth Plant Science Center, St. Louis, MO, (2)University of Illinois, Urbana, IL, (3)University of Nebraska-Lincoln, Lincoln, NE, (4)Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO

C_4 photosynthesis evolved multiple times independently in the angiosperms and over 22 times in the grasses suggesting that the ancestral genomes were primed for the acquisition of C_4 traits. However, the identification of the genes and networks associated with the transition from C_3 to C_4 have to date remained elusive. To accelerate the pace of gene discovery in the grasses, we have exploited the close phylogentic relationship of the panicoid grasses and deep genomic sequencing to identify signatures of adaptive evolution. Importantly, This approach does not rely on any *a priori* knowledge of the genes that contribute to biochemical or anatomical innovations associated with C_4 photosynthesis. Enabled by the collinearity of grass genomes and the recently sequenced genome of *Dichanthelium oligosanthes*, this analysis identified both known and novel components that likely underlie C_4 traits, including traits that are shared among all C_4 panicoids and those that are specific to one lineage.

W376: Functional Genomics of C_4 and CAM photosynthesis

Comparative Evolution of Crassulacean Acid Metabolism (CAM)

Xiaohan Yang[1], Rongbin Hu[1], John C. Cushman[2], Jerry Jenkins[3], Jeremy Schmutz[4], Gerald A. Tuskan[1], Hengfu Yin[5], Shengqiang Shu[6], Daniel S. Rokhsar[6], David M. Goodstein[6], Uffe Hellsten[6], Anne M. Borland[7], Ray Ming[8], Robert VanBuren[9], Ching Man Wai[8], Haibao Tang[10], Paul E. Abraham[1], Hao-Bo Guo[11], Rebecca L. Albion[2], Travis Michael Garcia[2], Jungmin Ha[2], Sung Don Lim[2], Jesse A. Mayer[2], Won Cheol Yim[2], Bernie Wone[12], Daniel Jacobson[1], Deborah Weighill[1], Xueqiong Li[13], Yanbin Yin[13], James Hartwell[14], Jim Leebens-Mack[15], Andrew Smith[16] and Klaus Winter[17], (1)Oak Ridge National Laboratory, Oak Ridge, TN, (2)Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV, (3)DOE Joint Genome Institute, Huntsville, AL, (4)Hudson Alpha, Huntsville, AL, (5)Chinese Academy of Forestry, Fuyang, China, (6)DOE Joint Genome Institute, Walnut Creek, CA, (7)Newcastle University, Newcastle Upon Tyne, United Kingdom, (8)University of Illinois at Urbana-Champaign, Urbana, IL, (9)Danforth Plant Science Center, St Louis, MO, (10)University of Arizona, Tucson, AZ, (11)University of Tennessee, Knoxville, TN, (12)University of South Dakota, Vermillion, SD, (13)Northern Illinois University, DeKalb, IL, (14)University of Liverpool, Liverpool, United Kingdom, (15)University of Georgia, Athens, GA, (16)Oxford University, Oxford, United Kingdom, (17)Smithsonian Tropical Research Institute, Balboa, Panama

Crassulacean acid metabolism (CAM) is a water-conserving photosynthetic pathway. CAM has evolved multiple times in a wide range of vascular plants. However, the molecular basis of CAM evolution is not well understood. The newly available CAM genomic resources provide an excellent opportunity for comparative analysis of CAM evolution in multiple plant lineages. In this study, we conducted genome-wide analysis of positive selection in protein-coding sequences in two CAM species, *Kalanchoë laxiflora* (representing an eudicotyledonous lineage),

and *Ananas comosus* (pineapple; representing a monocotyledonous lineage) in comparison with C_3 plant species. We compared day-night courses of gene expression between CAM and C_3 plant genes for thousands of orthologous groups. We also performed an in-depth analysis of selected gene families relevant to CAM physiology. Our results highlight the role of positive selection and diel re-programming of gene expression in CAM evolution. This study provides important novel insights into the molecular mechanism underlying CAM evolution.

W377: Functional Genomics of C_4 and CAM photosynthesis

Metabolic Flux Analysis and the Characterization of *dct2* in Maize

Douglas K Allen, USDA ARS, St. Louis, MO

W378: Functional Genomics of C_4 and CAM photosynthesis

Sequencing of the Pineapple Genome as a Model CAM species

Ray Ming[1,2], Robert VanBuren[3], Ching Man Wai[1], Haibao Tang[4], Michael Schatz[5], John E. Bowers[6], David Sankoff[7], Romain Guyot[8], Margaret R. Woodhouse[9], Yann-rong Lin[10], Ramanjulu Sunkar[11], Jim Leebens-Mack[12], Todd C. Mockler[13], Jeffrey L. Bennetzen[14], Andrew H. Paterson[15], Xinguang Zhu[16], Xiaohan Yang[17], Andrew Smith[18], John C. Cushman[19], Robert E. Paull[20] and Qingyi Yu[21], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)Fujian Agriculture and Forestry University, Fuzhou, China, (3)Danforth Plant Science Center, St Louis, MO, (4)University of Arizona, Tucson, AZ, (5)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (6)Department of Plant Biology, University of Georgia, Athens, GA, (7)University of Ottawa, Ottawa, ON, Canada, (8)IRD UMR DIADE, Montpellier cedex 5, France, (9)University of California, Berkeley, Berkeley, CA, CA, (10)Department of Agronomy, National Taiwan University, Taipei, Taiwan, (11)Oklahoma State University, Stillwater, OK, (12)University of Georgia, Athens, GA, (13)Donald Danforth Plant Science Center, Saint Louis, MO, (14)Dept. of Genetics, University of Georgia, Athens, GA, (15)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, (16)CAS-MPG Partner Institute for Computational Biology,Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China, (17)Oak Ridge National Laboratory, Oak Ridge, TN, (18)Oxford University, Oxford, United Kingdom, (19)Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV, (20)University of Hawaii at Manoa, Honolulu, HI, (21)Texas A&M AgriLife Research, Dallas, TX

Pineapple is the most economically valuable crop possessing crassulacean acid metabolism (CAM), a photosynthetic carbon assimilation pathway with high water use efficiency, and the second most important tropical fruit. We sequenced the genomes of pineapple varieties 'F153' and 'MD2', and a wild pineapple relative A. bracteatus accession CB5. The pineapple genome has one fewer ancient whole genome duplications than sequenced grass genomes and a conserved karyotype with seven pre rho duplication chromosomes. The pineapple lineage has transitioned from C3 photosynthesis to CAM with CAM-related genes exhibiting a diel expression pattern in photosynthetic tissues using beta-carbonic anhydrase(βCA) for initial capture of CO2. Promoter regions of all three βCA genes contain a CCA1 binding site that can bind circadian core oscillators. CAM pathway genes were enriched with cis-regulatory elements including the morning (CCACAC) and evening (AAAATATC) elements associated with regulation of circadian-clock genes, providing the first cis-regulatory link between CAM and the circadian clock regulation. Gene-interaction network analysis revealed both activation and repression of regulatory elements that control key enzymes in CAM photosynthesis. Pineapple CAM photosynthesis evolved by reconfiguration of pathways in C3 plants through regulatory neofunctionalization of preexisting genes and not acquisition of neofunctionalized genes via whole genome or tandem gene duplication.

W379: Fungal Genomics

Genome Sequencing-Assisted Identification of a Novel Type Pathway-specific Regulator and Dynamic Genome Environments of Solanapyrone Biosynthesis Gene Cluster in the fungus *Ascochyta rabiei*

Wonyong Kim[1], Jeong-Jin Park[1], David R. Gang[2], Tobin Peever[1] and **Weidong Chen**[3], (1)Washington State University, Pullman, WA, (2)Institute of Biological Chemistry, Washington State University, Pullman, WA, (3)USDA-ARS Washington State University, Pullman, WA

The fungal pathogen *Ascochyta rabiei* and some other unrelated fungi produce secondary metabolite solanapyrones or their derivatives. Secondary metabolite genes are often clustered together and situated in particular genomic regions like the subtelomere that can facilitate niche adaptation in fungi. Solanapyrones are toxic secondary metabolites produced by fungi occupying different ecological niches. Full genome sequencing of the chickpea pathogen *Ascochyta rabiei* revealed a solanapyrone biosynthesis gene cluster embedded in an AT-rich region proximal to a telomere end and surrounded by *Tc1/Mariner*-type transposable elements. The highly AT-rich environment of the solanapyrone cluster is likely the product of repeat-induced point mutations. Several secondary metabolism-related genes are found in the flanking regions of the solanapyrone cluster. Although the solanapyrone cluster appears to be resistant to repeat-induced point mutations, a *P450* monooxygenase gene adjacent to the cluster has been degraded by such mutations. Among the six solanapyrone cluster genes (*sol1-sol6*), *sol4* encodes a novel type of Zn(II)2Cys6 zinc cluster transcription factor. Deletion of *sol4* resulted in complete loss of solanapyrone production, but lack of solanapyrones did not affect growth, sporulation or virulence. Gene expression studies with the *sol4*-deletion and *sol4*-overexpression mutants delimited the boundaries of the solanapyrone gene cluster and revealed that *sol4* is likely a specific regulator for solanapyrone biosynthesis, and appears to be necessary and sufficient for induction of the solanapyrone cluster genes. Despite the dynamic surrounding genomic regions, the solanapyrone gene cluster has maintained its integrity, suggesting important roles of solanapyrones in fungal biology.

W380: Fungal Genomics

Comparative and Functional Genomics Analysis of the Magnaporthales

Laura Okagaki, William Sharpee, Josh Sailsbery, Alex Eyre, Titus John, Brent Clay, Yeonyee Oh and **Ralph A. Dean**, Center for Integrated Fungal Research, Raleigh, NC

The order Magnaporthales includes fungi of great economic importance that cause disease in cereal and turf grasses: *Magnaporthe oryzae* (rice blast), *Gaeumannomyces graminis* var. *tritici* (take-all disease), and *Magnaporthe poae* (summer patch disease). A genome-scale comparative

study was conducted across 74 fungal genomes. Gene clusters involved in transcriptional regulation and enzymatic activities were highly represented in order specific clusters and the species specific genes, suggesting that such proteins may be more plastic and may contribute to speciation. No correlations between diversifying or purifying selection and distance to repetitive elements or an increased rate of evolution in secreted and small secreted proteins were observed. Thus no evidence was found to suggest multi-speed genome evolution or that proximity to repetitive elements plays a role in diversification of genes. Functional analysis was conducted on candidate effectors (small secreted proteins) identified from comparative genome analysis of 40 *M. oryzae* strains, transcriptome and proteome data. Through a forward genetics screen, 11 suppressors of plant cell death (SPD) were found to inhibit the host cell death reaction induced by the BAX and/or the NPP1 genes within *N. benthamiana*. Four were previously identified as either essential for pathogenicity of *M. oryzae*, secreted into the plant during disease development, or homologues of other characterized suppressors. The others are novel and all remain to be fully characterized.

W381: Fungal Genomics

Comparative Genomics of Downy Mildews

Richard Michelmore, Genome Center, University of California, Davis, CA

Downy mildews (DMs) are oomycete pathogens that cause disease on a wide range of crop plants. Individual species have narrow host ranges and exhibit high degrees of host specialization. We utilized high throughput sequencing to generate *de novo* genome assemblies of geographically and temporally separated isolates of *Bremia lactucae* (lettuce DM), *Peronospora effusa* (spinach DM), *P. schachtii* (chard DM), *P. tabacina* (tobacco DM), *Peronosclerospora sorghi* (sorghum and maize DM), and *Sclerospora graminicola* (pearl millet DM) in collaboration with experts for each of these pathogens. These genome assemblies were highly syntenic but varied greatly in size (50 - ~320 Mb), gene content (12 – 20 K), repeat content (15 – ~60%), and repertoire of genes encoding effectors and other pathogenicity-related proteins. No families of RXLR-containing effectors are common across all DMs; however, RXLR families overlap greatly between isolates within a species. DM is the most important disease of lettuce in California and worldwide; *B. lactucae* is highly variable and can rapidly change to overcome resistant cultivars. We have monitored changes in virulence phenotype of *B. lactucae* in California since 1982. The appearance of the B_1 mating type may explain the recent increase in novel virulence phenotypes and may require a change in strategy for resistance gene deployment. Genomic sequence analysis of representative isolates collected in California and Europe over three decades provides detailed information of variation of *B. lactucae*.

W382: Fungal Genomics

A Comparative Gene Expression Analysis to Investigate Mechanisms of Fungal Wood-Decay

Melania Figueroa[1], Jiwei Zhang[1], Katarina Sweeney[1], Gerald Presley[1], Kenneth E. Hammel[2] and Jonathan Schilling[1], (1)University of Minnesota, St. Paul, MN, (2)USDA Forest Products Laboratory, Madison, WI

Brown rot and white rot fungi have evolved distinct mechanisms to colonize wooden structures and overcome lignin as the barrier to access sugars. In contrast to white rot fungi, which degrade lignin as part of their wood decay strategy, brown rot fungi deconstruct cellulose and hemicelluloses in very efficient fashion by only modifying this polymer. Understanding the roots of brown rot biodegradation of wood has important biotechnological and environmental applications such the production of antifungal wood preservatives and biofuels, as well as the conservation of ecosystems. Despite the significance of this biological process, solving the genetic and biochemical basis controlling wood decay has proven to be a challenge. The spatial complexity of the wood biodegradation process is one the limiting factors that often hampers experimental approaches. To alleviate this gap in knowledge, this project maps and integrates gene expression over networks of wood degrading fungal hyphae *in planta*; thus, providing resolution at selected mycelial regions exhibiting different metabolic properties. A comparative transcriptomics approach including representative brown rot and white rot fungal species has been set up to infer differences in the fungal secretome. Such differences are expected to explain the metabolic advantages harnessed by brown rot fungi and the evolution of these organisms.

W383: Fungal Genomics

Genomics of Anaerobic Cellulose-Degrading Fungal Symbionts of the Herbivore Gut

Alan Kuo[1], Stephen Mondo[1], Asaf Salamov[1], Kurt LaButti[1], John Henske[2], Kevin Solomon[2], Sean Gilmore[2], Michelle O'Malley[2] and Igor V. Grigoriev[1], (1)US Department of Energy Joint Genome Institute, Walnut Creek, CA, (2)UC Santa Barbara Dept. of Chemical Engineering, Santa Barbara, CA

As highly proficient degraders of cellulose, the Neocallimastigomycota fungi are key members of the microbiota of large mammalian and reptilian herbivores. Their remarkable cellulolytic capabilities have great potential for use in biofuel processing. They differ from other fungi 1) in possessing an extracellular cellulose-degrading complex called a fungal cellulosome, 2) in being obligate anaerobes, with hydrogenosomes instead of mitochondria, and 3) in developing flagella in certain life stages, placing them near the base of fungal phylogenies. All these distinctive characteristics touch on areas of intense interest to the US Dept. of Energy (DOE) Joint Genome Institute (JGI), including bioenergy, biotechnology, carbon cycling, fungal evolution, and symbiosis. The genomes of these interesting and important fungi are very AT-rich and highly repetitive, posing challenges to their genomic investigation. We have deployed PacBio long-read technology to overcome these challenges and to sequence, assemble, and annotate 3 Neocallimastigomycota genomes. We found modest numbers of genes (11-13k per genome) and gene families (< 1k clade-specific clusters), and confirmed low GC-content (< 25%), high repeat content (> 50%), absence of many aerobiosis-related genes (> 30 Core Eukaryotic Genes), and near-basal location in the fungal tree (branching with flagellated but aerobic Chytridiomycota). We are now using the genome to complement recent progress in elucidating the cellulosome using non-genome-based biochemical, proteomic, and transcriptomic methods.

W384: Fungal Genomics

Stage-Specific A-to-I RNA Editing in the Wheat Scab Fungus *Fusarium graminearum*

Huiquan Liu, Northwest A & F University, Yangling, China and **Jin-Rong Xu**, Purdue University, West Lafayette, IN

Yeasts and filamentous fungi do not have ADAR orthologs and are believed to lack A-to-I RNA editing, which is the most prevalent editing of mRNA in animals. However, during this study with the *PUK1* pseudo-kinase gene important for sexual reproduction in *Fusarium graminearum*,

we found that two tandem stop codons UA^{1831}G UA^{1834}G in its kinase domain were changed to UG^{1831}G UG^{1834}G by RNA editing in perithecia. To confirm A-to-I editing of *PUK1* transcripts, strand-specific RNA-Seq data were generated with RNA isolated from conidia, hyphae, and perithecia. *PUK1* transcripts were almost specifically expressed in perithecia and 90% of them were edited to UG^{1831}G UG^{1834}G. Genome-wide analysis identified 27,301 perithecium-specific A-to-I editing sites. Unlike those in animals, 70.5% of A-to-I editing sites in *F. graminearum* occur in coding regions and over two-thirds of them result in amino acid changes, including editing of 45 *PUK1*-like pseudogenes with stop codons in ORFs. *PUK1* orthologs and a number of other genes also displayed stage-specific expression and editing in *Neurospora crassa* and *F. verticillioides*. Furthermore, *F. graminearum* differs from animals in the sequence-preference and structure-selectivity of A-to-I editing sites. Whereas As embedded in RNA stems are targeted by ADARs, RNA editing in *F. graminearum* preferentially targets As in hairpin loops, which is similar to the anticodon loop of tRNA targeted by ADATs. Overall, our results showed that A-to-I editing occurs specifically during sexual reproduction and mainly in the coding regions in filamentous ascomycetes, involving adenine deamination mechanisms distinct from metazoan ADARs.

W385: Galaxy for SNP and Variant Data Analysis

Galaxy for SNP and Variant Data Analysis

Dave Clements, Johns Hopkins University, Eugene, OR

Galaxy is a data integration and analysis platform for life science research (http://galaxyproject.org). This workshop briefly introduces the Galaxy platform and then walks through a multi-step, SNP and variant calling analysis exercise, starting with quality control. We will take advantage of Galaxy's rich tool set and visualization capabilities to do this.

The workshop will also provide a brief overview of the Galaxy Project, and several ways in which the Galaxy platform is available to researchers.

Galaxy is an open-source, web-based platform, and there are over 60 publicly accessible servers around the world. Galaxy can also be installed locally, or on the cloud.

Galaxy enables bench scientists to create and experiment with sophisticated and *reproducible* data analysis. If you are new to next generation sequencing data analysis, or if you are trying to find a better way to manage and perform your analyses, then this workshop will be of interest to you.

W386: Gene Expression Analysis

Introduction

Gregory D. May, Pioneer HiBred, Johnston, IA

W387: Gene Expression Analysis

Relationships Between Nucleosome Positioning and Gene Expression

Shin-Han Shiu, Michigan State University, Lansing, MI

W388: Gene Expression Analysis

Nuclear Transcriptome Analyses of Specific Cell Types in the Early Plant Embryo

Daniel Slane, Max Planck Institute for Developmental Biology, Tuebingen, Germany

Organismal multicellularity goes along with the determination and specialization of single cells and ultimately tissue-types, which is in turn mainly dependent on the differential regulation of gene expression. To investigate the variation between specific transcriptomes, cells or even entire tissues of interest have to be isolated from the organism under study. In contrast to the inaccessible cell types of the early *Arabidopsis thaliana* embryos, the majority of whole-genome expression studies were carried out with total RNA from accessible root or shoot tissue. Nuclear RNA has been neglected for a long time as not being representative for transcriptomic studies, but there is accumulating evidence for the informative value of nuclear RNA. We recently described the generation, quality assessment and analysis of nuclear transcriptomic data from *Arabidopsis* embryos in comparison with total RNA transcriptomic data of comparable developmental stages. We used fluorescence-activated nuclear sorting (FANS) in combination with standard DNA microarrays to generate expression profiles and validated our datasets of differentially expressed candidate genes, proving the usefulness and applicability of this method for virtually any tissue type. Currently we are extending this technique to additional cell types of the early *Arabidopsis* embryo with a special focus on the epidermal (outer) cell layer. For this, we make use of RNA sequencing to detect even small transcriptional or RNA accumulation differences and to cover all transcripts expressed at a given time-point.

W389: Gene Expression Analysis

Single Cell Analysis of Cre-based Activation of Global Transcription

David W. Galbraith[1], Partha Samadder[2], Tom Doetschman[3], Ning Weng[2] and Ronald Heimark[4], (1)BIO5 Institute & School of Plant Sciences, University of Arizona, Tucson, AZ, (2)BIO5 Institute, University of Arizona, Tucson, AZ, (3)Department of Cellular and Molecular Medicine and BIO5 Institute, University of Arizona, Tucson, AZ, (4)Department of Surgery, University of Arizona, Tucson, AZ

The organs of eukaryotic organisms in general comprise a complex interspersion of different cell types, whose different molecular activities, and corresponding cellular states, cooperate during development to produce the final, functional organ. Dysfunction of organs in disease states, particularly oncogenesis, involves alterations in the genetic and epigenetic state of a minor subset of cells. It therefore is hard to detect early molecular indicators of disease states within the overwhelming background of normal cells. We have generated a number of transgenic mouse lines expressing a nuclearly-localized version of the Green Fluorescent Protein (GFP), in which production of a chimeric histone 2B-GFP protein is under the control of a constitutively-active, actin-derived promoter, separated by a Floxed-STOP sequence. In the presence of Cre recombinase, within F1 progeny, excision of the STOP sequence activates transcription and is identified by the emergence of green fluorescent nuclei. We describe the characterization of these lines using a combination of microscopic imaging, flow cytometry, and Reverse-Transcription

polymerase chain reaction of transcripts within single sorted nuclei. We describe strategies for employing these lines in the transcriptomic analysis of oncogenesis.

W390: Gene Expression Analysis

Characterizing Neuronal Diversity in the Adult Human Brain by Single-Nucleus RNA Sequencing

Blue B. Lake, University of California San Diego, La Jolla, CA

Understanding the human brain requires translating basic cellular physiology and neuroanatomy into an extensive interconnected network of ~100 billion neurons. Integral to this is the identification of neuronal subtypes based on functional gene expression profiles. However, characterizing the transcriptome of individual human adult neurons has been hampered by the difficulty in applying existing single cell RNA sequencing methodologies to the readily available repositories of postmortem tissues. Major challenges remain in the isolation of neurons that are intact, uncontaminated by glia, or which have minimal RNA degradation. To circumvent these issues, we have adapted the Fluidigm C1 microfluidics system to amplify total RNA from single neuronal nuclei for sequencing. Using this method, we have produced 3084 quality-filtered data sets from individual neuronal nuclei across six anatomically distinct regions of the prefrontal cortex. This sampling depth permitted unbiased classification of 16 distinct neuronal subtypes encompassing all of the cortical layers and revealed heterogeneity in expression profiles between functionally distinct brain regions that would be masked in bulk tissue studies. As such, we demonstrate a robust and scalable method for detection and categorization of single cell subtypes using difficult or limiting tissues through examination of their nuclear transcriptome.

W391: Gene Expression Analysis

TBGlobal Analysis of the RNA Secondary Structure and RNA-Protein Interaction Landscapes of Plants

Brian D. Gregory, University of Pennsylvania, Philadelphia, PA

At the heart of post-transcriptional regulatory pathways in eukaryotes are *cis*- and *trans*-acting features and factors including RNA secondary structure as well as RNA-binding proteins (RBPs) and their recognition sites on target RNAs. This is especially evident for RNA molecules whose functionality, maturation, and regulation requires formation of correct secondary structure and RNA-protein interactions. However, the global influence of these features on plant gene expression is still largely unclear. We have recently developed a high-throughput sequencing based approach that allows a simultaneous view of the RNA secondary structure and RNA-protein interaction site landscapes transcriptome-wide in eukaryotes. We have used this approach on multiple plant species and during their responses to various conditions and treatments. Our most recent findings from these studies will be presented.

W392: Gene Expression Analysis

An Extreme Metabolism: Iso-Seq analysis of the Ruby-Throated Hummingbird Transcriptome

Winston Timp, Johns Hopkins University Biomedical Engineering, Baltimore, MD

There are many features of natural and evolutionary history, morphology, and physiology that distinguish our study species, the ruby-throated hummingbird (***Archilochus colubris)***. They sustain the highest metabolic rates among all vertebrates; even while engaging in an annual migratory journey between Eastern North America and Central America. Notably, hummingbirds can switch rapidly (20-30 minutes) between a fuel of lipids to newly ingested sugars.

This remarkable metabolism is supported by enzymes which operate at the extreme limit of catalytic efficiency. Understanding the molecular basis of enzymatic action will provide a foundation enabling rational engineering of metabolic circuits in other systems.

To do this, we generated a de novo transcriptome of the hummingbird liver using PacBio IsoSeq, yielding a total of 8.6Gb of sequencing data, or 2.6M reads from 4 different size fractions. We analyzed the data with the SMRTAnalysis IsoSeq pipeline, including classification of reads, clustering of isoforms (ICE) followed by error-correction (Quiver). IsoSeq data is unique in that it provides full length mRNA transcripts, giving clear insight into coding sequences for novel protein products.

We used BLAST+ to search for human and chicken gene orthologs in our polished transcriptome. We also aligned our transcriptome against the *Calypte* draft genome where possible. Our characterization of the resulting protein coding sequences provides clues into how the hummingbird achieves its extreme metabolism.

W393: Gene Introgression

Germplasm Enhancement and Chromosome Remodeling in Wheat Wide >Hybridization

Fangpu Han, Chinese Academy of Sciences, Beijing, ChinaQinghua Shi, Xiang Guo, Jing Wang, Long Wang, Yanlin Hou, Zhenling Lv, Shulan Fu, Jing Zhang and Fangpu Han

Plant breeding may lead to a narrowing of genetic diversity of cultivated crops, thereby affecting sustained selection gains in crop improvement. Germplasm enhancement is an important aspect of wheat genetics and breeding. Thinopyrum elongatum and Thinopyrum intermedium, the wild relatives of wheat, have been suggested as a potentially novel source of resistance to several major wheat diseases including Fusarium Head Blight (FHB). A series of wheat (cv. Chinese Spring, CS)-Th. elongatum addition, substitution and ditelosomic lines were assessed for resistance to FHB. The results indicated that the lines containing chromosome 7E of Th. elongatum gave a high level of resistance to FHB; the infection did not spread beyond the inoculated floret. Furthermore, it was determined that the novel resistance gene(s) of 7E was located on the short arm (7ES) based on a difference in FHB resistance between the two 7E ditelosomic lines. Th. ponticum, Th. intermedium and tetraploid Th. elongatum contained useful and potential genes for wheat improvement. Amphiploids and partial amphiploids were released from the hybrids between wheat and Thinopyrum, and their genome chromosomal constitution was revealed by using GISH and multicolor GISH. Newly amphidiploids were obtained between wheat and rye, new 1B/1R translocation lines have been released and their centromere structure and functional analysis revealed commercial wheat varieties contained fusion centromere in this translocated chromosome.

W394: Gene Introgression

Germplasm Enhancement and Chromosome Remodeling in Wheat Wide Hybridization

Fangpu Han, Chinese Academy of Sciences, Beijing, China

W395: Gene Introgression

Trait Development *via* Mutagenesis in Tomato and Pepper

Daryl J. Somers[1], Travis W. Banks[1], David Liscombe[1], Anissa Poleatewich[1], Keiko Yoshioka[2] and Jas Singh[3], (1)Vineland Research and Innovation Centre, Vineland Station, ON, Canada, (2)University of Toronto, Toronto, ON, Canada, (3)Agriculture and Agri-Food Canada, Ottawa, ON, Canada

Advancements in genomics technologies, particularly genotyping and DNA sequencing, have led to a resurgence in using a reverse genetic approach to trait development. More specifically, researchers are using mutagenesis combined with high throughput DNA sequencing to develop novel alleles in genes that control biotic / abiotic stress and quality improvement. Over the past 4 years, we have developed methods and EMS mutagenized populations of petunia, tomato, pepper, cucumber and soybean that form a platform for trait development and functional genomics studies. Our approach includes using high resolution DNA melting for genotyping and Illumina-based sequencing coupled with informatics to discover and apply new alleles in plant breeding. The tomato (4,600 lines), pepper (3,400 lines) and soybean (4,300 lines) populations were determined to have mutation densities of 1 SNP/54 bp, 1 SNP/72 bp and 1 SNP / 40 bp respectively when the whole population is considered. We are pursuing trait development in tomato and pepper to improve disease resistance and quality characteristics such as flavour. The presentation will describe our rationale and process to generate both the genomics and biological resources needed for trait development and look to the future for emerging strategies.

W396: Gene Introgression

High-Density SNP Genotyping Array for Hexaploid Wheat and its Secondary and Tertiary Gene Pool

Mark Owen Winfield, University of Bristol, Bristol, United Kingdom

W397: Gene Introgression

Exploitation of Interspecific Diversity in Wheat

Julie King, The University of Nottingham, Leicestershire, United Kingdom

Due to modern breeding practises relatively little genetic variation is available in modern wheat varieties for breeders to develop superior adapted genotypes with increased yield potential and tolerance to abiotic and biotic stresses.

The wild relatives of wheat provide a vast and largely untapped reservoir of genetic variation (for traits such as tolerance to abiotic and biotic stresses, biomass, yield and photosynthetic potential).

Our primary objective is to transfer small, alien chromosome segments, carrying target genes but lacking deleterious genes, into hexaploid wheat quickly and efficiently. This is being achieved via exploitation of new marker technology, comparative mapping, exploitation of the genome sequence of model plant species and Next Gen Sequencing technology platforms to detect and characterise wheat/wild relative recombinants. Our initial results have shown very large numbers of wheat/wild relative introgressions, far beyond what we would have predicted from previous research, i.e. to date over a thousand potential introgressions from six different species. We believe that the strategy we have employed, accompanied with a new Affymetrix wild relative SNP array and genomic *in situ* hybridization techniques, provide a step change in our ability to transfer genetic variation into wheat from its distant relatives.

W398: Gene Introgression

Does Ph1 in Wheat Reveal the Universal Regulator of Pairing and Recombination?

Graham Moore, John Innes Centre, Norwich, United Kingdom

Despite possessing multiple sets of related chromosomes, hexaploid (bread) and tetraploid (pasta) wheat, both behave as diploids at meiosis. The *Ph1* locus in wheat ensures correct chromosome pairing and recombination between the related chromosomes in wheat. We have shown that: *Ph1* is defined to a kinase locus carrying a segment of heterochromatin; *Ph1* alters meiotic phosphorylation levels; *Ph1* promotes correct pairing of centromeres; *Ph1* promotes correct pairing of chromosome arms; *Ph1* regulates the conversion of MLH1 sites to crossover; and the ratio of MLH1 sites to crossover is closer to 1:1 in healthy wheat hybrids. Essentially *Ph1* has two distinct effects, regulating when the chromosomes become competent to pair during the telomere bouquet, and the ability of Double Holliday Junctions formed between diverged chromosomes to resolve as crossovers at the MLH1 complex stage. Consistent with this, initial studies indicate that key factors involved in initiating double strand break repair early in meiosis, and involved with the MLH1 complex, show altered phosphorylation with and without *Ph1*. Interestingly a recent C elegans study has also concluded that meiotic phosphorylation levels: promotes correct pairing of centromeres; promotes correct pairing of chromosome arms; regulates the conversion of MLH1 sites to crossovers; and at increased phosphorylation levels, ratio of MLH1 sites to crossover is closer to 1:1 in healthy worms. Therefore phosphorylation may be the universal regulator of correct meiotic chromosome pairing and recombination.

W399: Gene Introgression

Development of Chromosome Specific Molecular Markers for *Thinopyrum elongatum* and their Applications in Gene Introgression

Jianmin Chen, Yangzhou University, Yangzhou, China

It is necessary to develop specific markers of *Th. elongatum* for useful genes to be transferred into wheat. The eighty nine chromosome specific molecular markers for 7E were developed by Specific Length Amplified Fragment Sequencing (SLAF-seq) with a successful rate of 65.9%, The thirteen specific markers to 1E~7E chromosome of *Th. elongatum* by Retrotransposon microsatellite amplified polymorphism (REMAP) and thirty one molecular markers specific to each E-chromosome of *Th. elongatum* were developed by Target region amplification polymorphism (TRAP) technologies. These chromosome-specific molecular markers were applied to detect chromosomes and /or fragments of *Th. elongatum* in wheat background.The results of this study have provided with important theoretical and practical basis for the study and utilization of *Th. elongatum* involved wheat germplasm.

The *Durum- Th. elongatum* addition lines were established based on combined techniques of cytology and molecular marker selection. The 2E, 7E chromosome disomic addition lines and, 4E,5E chromosome monsomic addition lines in F_4 progeny were identified by specific molecular markers and GISH. The wheat-*Th. elongatum* 7E chromosome long arm translocation was obtained using 7E chromosome specific molecular markers and genomic *in situ* hybridization (GISH).The result of FHB reaction evaluation indicated that the 7EL translocation lines had a higher level of resistance than that of the 7ES.
The progenies of synthetized hexaploid (AABBEE) crossed with triticale (AABBRR) to were analyzed by specific molecular marker and GISH for introgression of *Th. elongatum* chromosomes. The results of GISH revealed that several triticale lines had E chromosomes and/or E-R chromosome translocation. Analyses of R and E chromosome-specific molecular marker confirmed the above results. This investigation leads to conclude that specific molecular marker can track each E and R genome chromosomes and /or fragments introgression.

W400: Genome annotation resources at the EBI

Intro

Sandra Orchard, EMBL-EBI, Hinxton, United Kingdom

The European Bioinformatics Institute
EMBL-EBI provides freely available data from life science experiments to users across the globe, performs basic research in computational biology and offers an extensive user training programme, supporting researchers in academia and industry. We house the world's largest collection of biological databases in a single institute and produce many of these in collaboration with other major institutes and research groups, including the NCBI, Swiss Institute of Bioinformatics and the Wellcome Trust Sanger Institute. We, and representatives from the group we collaborate with, will present some of these resources, including a brief overview of the data they contain and how to access and use this valuable information. Presenters will then be available for detailed questions both after the workshop and at the EBI stand in the exhibition hall.

W401: Genome annotation resources at the EBI

Browsing Genes and Genomes with Ensembl and Ensembl Genomes

Helen Sparrow, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Browsing Genes and Genomes with Ensembl and Ensembl Genomes will be presented by Helen Sparrow, Ensembl Outreach Officer. The talk will include an introduction to Ensembl browsers, key views in browsing genomes, and tools for accessing genomic data and analysing your own, including BioMart and the VEP.
Ensembl (www.ensembl.org) provides an interface and an infrastructure for accessing genomic information covering over 70 vertebrate species, including cow, pig, sheep, and chicken. Its sister project, Ensembl Genomes (http://www.ensemblgenomes.org), consists of five sub-portals (bacteria, protists, fungi, plants, and invertebrate metazoa) which contain data for over 700 eukaryotic (including wheat, barley, tomato and the 12 OGE species) and almost 30,000 prokaryotic genomes.
All species in Ensembl and Ensembl Genomes have gene annotation (based on biological evidence) and comparative genomics analyses within the taxa (excluding bacteria). For many of these genomes, we also provide annotation of sequence variants, such as SNPs and CNVs. Access to these data are via our browser websites, BioMart (for protists, fungi, plants, and animals), FTP, Perl APIs, REST API, and MySQL queries. The Variant Effect Predictor (VEP) is a powerful tool that allows analysis of your own sequence variants and is available for all species in Ensembl and Ensembl Genomes and can even be used for species not in Ensembl. Highlights of the past year include; updated phenotype data for sheep and chicken, pairwise alignments for more than 50 plant genomes, and new variation data for wheat and barley.

W402: Genome annotation resources at the EBI

Community Manual Genome Annotation and Vega

Jane Loveland and Jennifer Harrow, Wellcome Trust Sanger Institute, Cambridge, United Kingdom

The Human and Vertebrate Analysis and Annotation (HAVANA) team at the Wellcome Trust Sanger Institute (WTSI) undertakes manual annotation of vertebrate genomic sequence. We have whole genomes for human, mouse, zebrafish, pig and rat and specific regions of interest, such as the Major Histocompatibility Complex (MHC), for selected organisms). Our manual annotation is publicly available from the Vertebrate Genome Annotation Database (VEGA) (www.vega.sanger.ac.uk). The HAVANA team produces the manual annotation that is integrated in the GENCODE reference gene set for human and mouse. This can be downloaded from gencodegenes.org and is available from the Ensembl and UCSC genome browsers.
The Otterlace/Zmap annotation tools that have been developed at the WTSI, are being used remotely by external collaborators for community annotation. We have held annotation workshops for several species, most recently rat, that have utilised these tools and generated high quality manual annotation. To date have annotated over 2200 genes chosen by the rat community. Gene clusters are particular targets for manual annotation due to the difficulty with their automated annotation. We recently completed the annotation of the MHC (RT1) region in rat.
Annotation is a continuous process and so between database updates we can release new annotation via a Vega update track for all of our whole genome species.
http://vega.sanger.ac.uk/info/data/frequent_update.html
If you have any queries please contact us at:
vega-helpdesk@sanger.ac.uk

W403: Genome annotation resources at the EBI

Introducing the Vertebrate Gene Nomenclature Committee (VGNC)

Susan Tweedie, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

Standardised gene nomenclature provides an essential resource for all researchers. However an ever-increasing number of vertebrate genomes are being sequenced and the data released into the public domain without any systematic annotation or gene naming. There are currently only six vertebrate model organisms with an official gene nomenclature group (mouse, rat, chicken, Anolis, Xenopus and zebrafish), all of which base their gene names on those approved by the HUGO Gene Nomenclature Committee (HGNC) for human genes. This presentation introduces the

Vertebrate Gene Nomenclature Committee (VGNC) – a parallel branch of the HGNC tasked with approving gene names and symbols across vertebrates.
Our naming strategy for each vertebrate species starts by identifying a high confidence set of genes with 1:1 human orthologs using our HCOP tool, which aggregates orthology predictions from many sources. These 1:1 orthologs will be named in a semi-automated manner, with the human gene nomenclature being transferred to the orthologous gene. Genes with non-consensus orthologs, members of complex gene families, pseudogenes and RNA genes will require additional manual curation.
Our prototype species for naming is chimpanzee and we have begun naming the 12,952 protein-coding chimp genes with a 1:1 human ortholog. During this process we are taking the opportunity to improve the consistency of our names and taking care to minimise transfer of species-specific information (e.g. molecular weights). This naming process will soon be expanded to other species, including dog and cow.
Further information and requests for individual vertebrate gene names and symbols can be made via: http://www.genenames.org.

W404: Genome annotation resources at the EBI

The European Variation Archive: A Central Resource for Genetic Variation Data

Gary Saunders, EMBL-EBI, Cambridge, United Kingdom

The European Variation Archive (EVA; www.ebi.ac.uk/eva) archives all types of variation data from all species, and currently more than 40% of all variants archived at EVA are from non-human organisms, mostly plant and livestock. We aim to provide a resource to the plant and animal community that can accession, archive and provide views of genetic variation data more quickly and in a more granular manner than any other resource of this nature, worldwide.
EVA predominantly archives genetic variation data in Variant Call Format (VCF) files, which is the community standard for describing and sharing variants. We shall present an overview of the ways in which EVA works with data submitters during the submission process to confirm that archived data is truly valid and associated with rich metadata to ensure that these data are of most benefit to the community.
At EVA, archived variants are normalized and annotated using a variety of standardized methods, including Ensembl's Variant Effect Predictor. We shall present the ways in which users can mine these data using filters on the website to construct both study-centric and global queries, filtering on any combination of species, methodology, variant type, phenotype, consequence or allele frequency and show how results from these queries can be downloaded in a variety of formats including VCF and CSV. Additionally, EVA provides a comprehensive RESTful web-service, to allow programmatic access, and hence the integration of these data with other resources such as Ensembl Variation and Uniprot, and this shall also be presented.

W405: Genome annotation resources at the EBI

Functional Genomics Data and Expression Look-up Tools: ArrayExpress and Expression Atlas

Amy Tang, EMBL-European Bioinformatics Institute, Hinxton, United Kingdom

Expression Atlas at EMBL-EBI contains pre-analyzed RNA-seq and expression microarray data for researchers to discover and query which genes are expressed in which tissues, cell types, developmental stages, and many other experimental conditions. Queries can either be in a baseline context, e.g. find genes expressed in the bovine kidney, or in a differential context, e.g. find genes that are up/downregulated in response to salt stress in rice. All datasets are manually curated to a high standard by our in-house curators and processed by our standardised statistical pipeline, both in consultation with experts around the world. As of October 2015, Expression Atlas consists of 2162 datasets, including 120 RNA-seq experiments. All data in Expression Atlas is free to browse, download, re-use, and is originated from the ArrayExpress archival database of functional genomics experiments at EMBL-EBI.
ArrayExpress experiments were either directly submitted by scientists working in diverse fields, or imported systematically from NCBI GEO weekly. As of October 2015, ArrayExpress consists of 60838 datasets (80% from microarrays) studying a wide variety of organisms, from cattle and chicken to barley and sorghum. Public ArrayExpress data sets are free for download, either from the website or programmatically.
Both databases employ EFO ontology-driven query expansion, enabling powerful searching across thousands of datasets.
Submission to ArrayExpress is a free service via a webform-based tool called Annotare. To facilitate peer review, accession numbers are generated usually within 15 minutes of submission, pre-published data sets can be kept private, and submitter's identity can be hidden for double-blind review.

W406: Genome annotation resources at the EBI

Expert Curation of Proteins in UniProtKB/Swiss-Prot

Damien Lieberherr and **Sylvain Poux**, 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. We recently started to collaborate with Araport, the Arabidopsis portal, and 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 manual annotation has been added to more than 4500 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.

W407: Genome annotation resources at the EBI

The Gene Ontology and Its Annotation Sets

Claire O'Donovan, European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom

The Gene Ontology (GO) project is a collaborative effort to address the need for consistent descriptions of gene products across databases and for usage by researchers worldwide to describe their data. The GO project has developed three structured ontologies that describe gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner. There are three separate aspects to this effort: first, the development and maintenance of the ontologies themselves; second, the annotation of gene products, which entails making associations between the ontologies and the genes and gene products in the collaborating databases; and third, the development of tools that facilitate the creation, maintenance and use of ontologies. In this presentation, I will address these three aspects but will focus on the process of expert manual curation and explain what is available for use by this community and how this community can contribute themselves.

W408: Genome annotation resources at the EBI

InterPro and Automatic Annotation of Non-Model Organism Proteomes

Sandra Orchard, EMBL-EBI, Hinxton, United Kingdom

InterPro, a tool for annotation transfer

InterPro provides functional analysis of proteins by classifying them into families and predicting domains and important sites. The database combines protein signatures from a number of member databases into a single searchable resource, capitalising on their individual strengths to produce a powerful integrated database and diagnostic tool. InterProScan is a sequence analysis application (nucleotide and protein sequences) that combines different protein signature recognition methods into one resource, widely used by genome sequencing groups to predict the function of the coding genes in their organism of interest. The tool is used at the EBI for the transfer of annotation from well studied proteins which have been manually annotated by UniprotKB/Swiss-Prot and Gene Ontology curators to experimentally uncharacterised gens and protein in non-model organisms. The mechanism and rules for annotation transfer will be described and briefly discussed.

W409: Genome annotation resources at the EBI

Training on Data, Tools and Resources for Life Scientists

Katrina Costa, EMBL-EBI, Cambridge, United Kingdom

There is a wealth of biological data being generated by laboratories worldwide, which is then deposited into publically available databases. This creates an increased need for bench biologists to gain the skills required to access, analyse and interpret the data.

At EMBL-EBI (the European Bioinformatics Institute) we develop and maintain a broad range of data resources that span the life sciences, and also provide various training opportunities for scientists to learn how to get the most from their biological data.

Many of our trainees are from a bench background and at PhD/Postdoctoral level, but whilst they have expertise in their area of scientific interest they often need help developing computational skills and a knowledge of the tools available to them. Our aim is not to produce a new generation of bioinformaticians, but to develop a group of scientists who are more confident users of data tools and resources.

This talk will introduce you to our training programme, which delivers training in three ways: on-site, off-site and online.

W410: Genome management and analysis with CoGe

Introduction to CoGe

Eric Lyons, University of Arizona, Tucson, AZ

With 25,000 genomes from 17,000 different organisms, CoGe is where many people are managing and analyzing their genomic data. This workshop will provide an overview and demonstration of CoGe's latest technologies including RNASeq processing, genome analysis, data management, APIs (Application Programming Interfaces), and integration with other bioinformatic resources. CoGe is available at http://genomevolution.org.

W411: Genome management and analysis with CoGe

Next-Gen Sequence Analysis in CoGe: Read Alignment, Expression Analysis, SNP Identification, and What's to Come

Matthew Bomhoff, University of Arizona, Tucson, AZ

Next-Gen Sequence Analysis performed by hand is a time-consuming process that requires specialized knowledge of computer hardware and operating systems, as well as an array of disparate bioinformatic software packages. CoGe's new Next-Gen Sequence Analysis toolkit dramatically simplifies the process by automating workflow generation through an easy-to-use web application. The user is guided through the process of uploading their FASTQ read data to CoGe (via desktop, FTP, or the iPlant Data Store) and provided with a menu of options for read alignment, expression analysis, and SNP identification. Many common third-party bioinformatic tools are available, such as TopHat, HISAT2 and GSNAP for read alignment, Cufflinks for expression analysis, and Platypus and SAMtools for SNP identification. Analysis results can be viewed in CoGe's genome browser, and optionally shared with other CoGe users. Future downstream analyses currently under development include diversity quantification, ChIP-seq analysis, and GWAS.

W412: Genome management and analysis with CoGe

APIs and Automated Workflows

Sean Davey, University of Arizona, Tucson, AZ

Automated workflows - using CoGe APIs to create and manage job workflows

CoGe's expanding application program interfaces (APIs) provide programatic access to the data and tools normally available via the CoGe web application user interface. An API is a way to submit a request (via an HTTP URL) and get some data back (in this case in JSON format).

CoGe has RESTful APIs for managing (creating, updating, deleting, retrieving) different classes of data, and APIs for submitting jobs and retrieving the results from the job runs. In this talk I will describe the

general mechanism for submitting API requests and present several examples showing their use. Some jobs can be used for batch submission or retrieval of data, while others are used for initiating workflows using various CoGe tools and for obtaining the results of those workflows. I will also discuss possible future expansion to the APIs which may provide access to additional data, new searching ability within CoGe, new tools and possibly novel workflows based on client interest.

W413: Genome management and analysis with CoGe

FractBias: Graphical and Integrated Tool for Assessing Fractionation Bias after Whole Genome Duplications

Blake L. Joyce[1], Sean Davey[1], Matthew Bomhoff[1], James C Schnable[2] and Eric Lyons[1], (1)University of Arizona, Tucson, AZ, (2)University of Nebraska-Lincoln, Lincoln, NE

Whole genome duplication and fractionation, the loss of duplicated genes, disrupt gene order within genomes over evolutionary time. FractBias is a new CoGe tool that assesses fractionation bias in a genome that has experienced a whole genome duplication event (query genome). The query genome is then compared to a close relative that did not experience that whole genome duplication (target genome). Syntenic regions are found using the SynMap tool, and then the percentage of retained genes on each chromosome of the query genome are graphed for every chromosome in the target genome. This permits the evaluation of bias among homeologous chromosomes, which has been demonstrated to be in an important driver shaping the evolution of genome structure. Ultimately, genes from each genome can be exported for annotation to investigate how genome duplication, fractionation bias, and gene functions interplay.

W414: Genome management and analysis with CoGe

High-Performance and Complex Data Visualization in Genomics

Asher K Haug-Baltzell and Eric Lyons, University of Arizona, Tucson, AZ

Due to recent advances in sequencing technologies and corresponding decreases in sequencing costs, biological researchers are awash in data. This surge of genome datasets, combined with the vast collection of already available data, is incredibly exciting for researchers. However, the scale of the available information and complex relational nature pose significant challenges, particularly when performing comparisons of many genomes and datasets. The Comparative Genomics platform CoGe is working towards developing novel visualizations that will allow for researchers to draw meaningful conclusions from such massive genomic datasets. With a current database of nearly 25,000 genomes for 17,000 species, a huge collection of experimental data, and underlying infrastructure for computational pipeline integration and execution, CoGe is uniquely positioned to begin investigating these new visualizations. Over the last year, CoGe has implemented two new data visualization tools: a high-performance dot-plot viewer and a three-dimensional three-genome synteny browser, both discussed here. These systems are important steps for increasing researchers' ability to visualize and interact with high-dimensionality data, both through modern web-based technologies and newly-available virtual reality environments.

W415: Genome management and analysis with CoGe

Integration of CoGe's Data Services in iPlant

Jeremy D. DeBarry, University of Arizona, Tucson, AZ

As the number and capabilities of computational research platforms expands, there is a growing need for interoperability. To accomplish research goals, researchers increasingly rely on seamless transfers of data, metadata, and experimental results between platforms. Without this, analyses involving multiple platforms are often prohibitive due to the time and effort spent moving and formatting data to meet needs not directly related to the research focus. The Comparative Genomics platform, CoGe, has coordinated with the iPlant Collaborative to enable researchers to search CoGe's catalog of nearly 23,000 public genome sequences from over 17,000 species, and import them directly into the iPlant Discovery Environment. From the Discovery Environment, researchers can harness powerful compute resources and choose from among hundreds of bioinformatics Apps. This allows users to seamlessly leverage the strengths of both CoGe and iPlant for genome analyses. The ongoing collaboration between CoGe and iPlant exemplifies the value added when interoperability is leveraged to create an ecosystem of resources that enable and empower researchers, accelerating discovery.

W416: Genome management and analysis with CoGe

fRANkenSeq

Allen Hubbard, University of Delaware, Newark, DE

W417: Genomic features and chromosome functionality

A Brief Introduction to the Workshop and My Research

Xiyin Wang, School of Life Sciences, North China University of Science and Technology, Tangshan, Hebei, China; University of Georgia, Athens, GA

The workshop would be a platform for scientists working to understand how genomic features affect the function of chromosomes. The eukaryotic genome is packed into chromosomes, which function regulation and interaction between DNA, RNA, and proteins. Chromosomes have two main functions: to ensure that the DNA is segregated equally to daughter nuclei at cell division, and to ensure that the integrity of the genome is maintained and accurately replicated in each cell cycle. The elements responsible for these functions are centromeres, telomeres and replication origins, respectively. Genome sequencing has been providing enormous materials to understand chromosome biology. From these precious materials, scientists are finding how genomic features, such as base composition, homoeologous or ectopic sequence similarity, distribution of tandem genes, recombination hotspot, etc, affect the transcription, expression, functional innovation, and concerted evolution of genes, and chromosomal rearrangement, chromosome number reduction, and even whole genome repatterning.

W418: Genomic features and chromosome functionality

Structural Evolutionary Dynamics in Allopolyploid Synthetic Oilseed Rape (*Brassica napus* L.)

Mathieu Rousseau-Gueutin[1], Jérôme Morice[1], Sylvie Nègre[1], Gwenn Trotoux[1], Cyril Falentin[1], Olivier Coriton[1], Virginie Huteau[1], Florian Kerbrat[1], Gwenaelle Deniot[1], Marie Gilet[1], Sonia Vautrin[2], Joelle Fourment[2], Maryse Lodé[1], Alexandre Pelé[1], Frederique Eber[1], Hélène Bergès[2] and Anne-Marie Chevre[1], (1)INRA, Le Rheu, France, (2)INRA - CNRGV, Castanet Tolosan, France

Allopolyploidy, which results from the merger and duplication of two divergent genomes, has played a major role in the evolution and diversification of flowering plants. This evolutionary success of allopolyploid species results from their higher structural and functional evolutionary dynamic compared to their diploid parents, which begins immediately after the allopolyploidization event. Presently, a model system to study this phenomenon is *Brassica napus* (oilseed rape) since its whole genome sequence has been sequenced recently. In order to detect the immediate impact of allopolyploidy on the structural evolutionary dynamic of *B. napus* (2n=4x=38), two synthetic *B. napus* populations were created. By comparing the genome structure of 30 synthetic *B. napus* individuals to their parental diploid genomes using a Illumina SNP 60k array, we were able to identify important deletions in each synthetic plant, which ranged from 0.1 to more than 20 Mb (whole chromosome arm). In order to determine the type of rearrangements involved (either deletion or non reciprocal translocation), BAC-FISH experiments (Fluorescence In Situ Hybridization) were performed using BACs specific to the rearranged regions. Compared to what was previously known, we showed statistically that some chromosomes and even one of the subgenomes were more prone to structural modifications. In addition, our findings showed that the *B. napus* genome may be deeply shuffled only two generations after the allopolyplodization event. In some individuals, these structural rearrangements may cause the loss of one homoeologous copy for few thousands of genes.

W419: Genomic features and chromosome functionality

The Population Genomic Structure of the Asian Cultivated Rice (*Oryza sativa* L.) I: Genome Sizes, Pan Genomes, and Structural Variation

Zhi-Kang Li, Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Beijing, China

As the staple food for half the world's population, rice (*Oryza sativa* L.) is also well known for its subspecific differentiation and rich within species diversity. As a major international effort to understand the population genomic structure of within *O. sativa* diversity, a core collection of 3,010 rice accessions from 89 countries were re-sequenced with an average sequencing depth of 14.9x. Based on SNP variation, the 3,010 rice accessions were classified into two major subspecies, 1,764 *Xian* (*Indica*) accessions, 800 *Geng* (*Japonica*) ones, 221 AUS accessions, 101 aromatic accessions, plus 124 admixtures. Deep analyses of this huge dataset revealed, for the first time, the three important aspects of the population genomic structure of the within *O. sativa* diversity: genome size (GS) variation, pan genomes (PG) and structural variation (SV). First, de novo assembly of the 3,010 rice genomes produced an average assembled GS of 288.2±21.5Mba and an estimated average GS of 375.2±21.3 Mbp, with a mean repetitive sequence (RS) content of 35.6±3.7%, which is highly correlated with the estimated GS. Secondly, in our PG analyses, a total of 17,893 protein-coding genes were predicted from the 320 Mbp novel sequences of 453 genomes of high quality. These novel genes plus those in the reference Nipponbare genome were merged into 24,070 gene families, of which 11,027 gene families were present in all rice accessions, forming the core PG of *O. sativa*, while the remaining 13,043 gene families comprise the distributed PG of rice. Using computer simulation, we were able to determine 1,127 undiscovered gene families in the distributed PG of *O. sativa*. Thus, the PG of *O. sativa* consists of 25,197 gene families, including 11,027 core gene families (43.8%) and some 14,170 (56.2%) distributed gene families. Classification of the 453 rice accessions based on the presence/absence of the PG gene families was very similar to that based on SNV with very few misclassifications. Thirdly, by directly comparing the contigs of each assembled genome with the Nipponbare and 9311 reference genomes, we discovered large numbers of large SVs. Of these SVs, translocations and deletions were predominant and accounted for most of the SVs detected, while duplications and inversions accounted for small portions of the detected SVs. The detected large SVs were not randomly distributed across the rice genome with many hotspots of SVs identified. Together, our results provided the most comprehensive picture on the population genomic structure of *O. sativa* and insights into the evolution and origin of this important species.

W420: Genomic features and chromosome functionality

Duplicate Gene Divergence by Changes in MicroRNA Binding Sites in Arabidopsis and Brassica

Keith Adams, University of British Columbia, Vancouver, BC, Canada

Gene duplication provides large numbers of new genes that can lead to the evolution of new functions. Duplicated genes can diverge by changes in sequences, expression patterns, and functions. MicroRNAs play an important role in the regulation of gene expression in many eukaryotes. After duplication, two paralogs may diverge in their microRNA binding sites, which might impact their expression and function. Little is known about conservation and divergence of microRNA binding sites in duplicated genes in plants. We analyzed microRNA binding sites in duplicated genes in Arabidopsis thaliana and Brassica rapa. We found that duplicates are more often targeted by microRNAs than singletons. The vast majority of duplicated genes in A. thaliana with microRNA binding sites show divergence in those sites between paralogs. Analysis of microRNA binding sites in genes derived from the ancient whole-genome triplication in B. rapa also revealed extensive divergence. Paralog pairs with divergent microRNA binding sites show more divergence in expression patterns compared with paralog pairs with the same microRNA binding sites in Arabidopsis. Close to half of the cases of binding site divergence are caused by microRNAs that are specific to the Arabidopsis genus, indicating evolutionarily recent gain of binding sites after target gene duplication. We also show rapid evolution of microRNA binding sites in a jacalin gene family. Our analyses reveal a dynamic process of changes in microRNA binding sites after gene duplication in Arabidopsis and highlight the role of microRNA regulation in the divergence and contrasting evolutionary fates of duplicated genes.

W421: Genomic features and chromosome functionality

Genomic Discovery through Chromosome Manipulation

Isabelle M. Henry[1], Ek Han Tan[1] and Luca Comai[2], (1)University of California, Davis, CA, (2)Plant Biology and Genome Center, UC Davis, Davis, CA

From conventional breeding to random mutagenesis and genome editing, approaches to functional genomics are diverse and rapidly improving in efficiency and precision. Here we will discuss examples of functional genomics through the manipulation of chromosomes and their respective advantages and disadvantages. In clonally propagated species, random large-scale insertions and deletions can easily be maintained and provide a

wide-range of phenotypic variability that can be harnessed both for gene discovery and crop improvement purposes. A different approach consists in manipulating centromere function to shape genomes. We will discuss the mechanisms underlying haploid induction through altered centromere function, as well as its different applications, including the potential use of centromere manipulation to rapidly introgress valuable traits from exotic germplasm while avoiding linkage drag of deleterious alleles.

W422: Genomic features and chromosome functionality

Asymmetrical Genome Evolution and its Impact on Trait Formation in Brassica Crops

Chaobo Tong[1], Xianhong Ge[2], Rod Snowdon[3], Boulos Chalhoub[4], Tae-Jin Yang[5], Zaiyun Li[2] and **Shengyi Liu**[1], (1)Oil Crops Research Institute of CAAS, Wuhan, China, (2)Huazhong Agricultural University, Wuhan, China, (3)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (4)URGV-INRA, Evry, France, (5)Seoul National University, Seoul, South Korea

Genome polyploidization has provided significant sources of genetic variation for plant adaptive evolution and new species formation. However, the way in which molecular evolution of polyploid genomes builds up genetic architecture underlying speciation is unclear and its impact imposed on trait formation involving duplicate genes is unknown. *Brassica* is an ideal model to address these questions and thereby we conducted comparative genome analysis of newly sequenced *Brassica oleracea, B. rapa and B. napus*.

We revealed multi-layered modes of asymmetrical subgenome evolution. These layers include: massive and asymmetrical subgenomic gene loss, asymmetrical variations in paralogous DNA sequences, expression differentiation of triplicated, α-duplicated and tandem duplicated genes, asymmetrical alternative splicing variants between the subgenomes (Liu et al. Nature Communications 5:3930, 2014) and asymmetrical homeologous exchanges (Chalhoub et al. Science, 2014), asymmetrical epigenomes and asymmetrical recombination between the genomes A and C in *B. napus*. The epigenomes include small RNA, DNA methylation and histone modification. Further, we used associated markers from a genome-wide association study of a large population to have revealed differences in detectable traits such as flowering time and oil content between syntenic regions of subgenomes A and C. These patterns provide new insight into genome evolution underlying speciation and trait formation and will underpin research in genetic improvement of these important crops.

W423: Genomic features and chromosome functionality

Features of RNA Polyadenylation Sites in Various Algae and Pathogenic Protists

Xiu-Qing Li, Agriculture and Agri-Food Canada, Fredericton, NB, Canada

Protists are very diverse in genome and include various species that are pathogens to crops and animals. For example, Phytophthora species, including the potato late blight pathogen, can cause severe diseases to crops. At the same time, various protists are algae that might be relatives to plants and are model species for research. Here I present the features of RNA polyadenylation sites of various protists. These features may provide insights about the potential relationship of these protists to plants and animals, facilitate genetic manipulation of these species, and help development of drugs or pesticides against some pathogens.

W424: Genomic Selection and Genome-Wide Association Studies

Identification of Candidate Causal Variants Underlying QTL in Dairy Cattle through GWAS and Bayesian Approach at the Sequence Level

Didier A. Boichard[1], Marie-Pierre Sanchez[1], Anne Barbat[1], Mekki Boussaha[1], Thierry Tribout[1], Rachel Lefebvre[1], Sébastien Fritz[2], Romain Saintilan[2], Chris Hoze[2], Armelle Govignon-Gion[1] and Pascal Croiseau[1], (1)INRA UMR1313 Animal Genetics and Integrative Biology, Jouy en Josas, France, (2)Allice, Genetics Team, Paris, France

Combining large genotyped and phenotyped resource populations with whole genome sequence information opens new avenues to unravel candidate mutations underlying the genetic variability of the phenotypic traits. Reference populations for genomic selection are very large (more than 100,000 dairy animals with phenotypes and genotypes in France) and increase rapidly. As shown by the 1000 bull genomes project (1682 sequenced bulls in 2015), the number of sequenced bovine genomes also increases very fast and already makes it possible to impute the reference populations up to the sequence level, although rare alleles still have limited imputation accuracy. These populations sequenced in silico are unique resources for GWAS studies. Several strategies can be applied: first, each population can be analyzed by simple GWAS one marker at a time; results for the same trait but from multiple breeds can be combined in meta-analyses; multi-markers approaches can be implemented on targeted regions (1-3 Mb) in order to reduce the effect of long range linkage disequilibrium; haplotype analyses can be used to filter out genetic variants with inconsistent effects; finally, functional annotations help selecting the best candidates and interpreting their likely effects. Examples from the PhenoFinLait project are given, showing that QTL involved in milk protein or fatty acids composition can be characterized with a high confidence up to the gene and, frequently, up to a very limited number of candidate mutations. The authors acknowledge the financial support from ANR and APIS-GENE, and the contribution of the 1000 bull genomes consortium.

W425: Genomic Selection and Genome-Wide Association Studies

Multi-Locus Methods for Genome-Wide Association Studies and Genomic Selection

Min Zhang and Dabao Zhang, Purdue University, West Lafayette, IN

With the development of modern high throughput genotyping technology, an increasingly large amount of genomic polymorphism data have been generated and therefore, genome-wide association studies are becoming a popular tool to identify associations between genomic variants and quantitative traits. Despite the consensus regarding the value of genome-wide association studies for improving our understanding of complex traits, there are well documented problems in the statistical analysis of these data, such as the much larger number of polymorphisms compared to the limited number of subjects, the linkage disequilibrium between polymorphisms, and the confounding issue introduced by population structure, among many others. To meet these challenges, we have developed a series of multi-locus methods, including penalized orthogonal components regression to handle continuous phenotype in association analysis, and generalized orthogonal components regression for case-control studies. Recently, similar methods were developed for mixed models to account for the family structure in family based genome-wide association studies. Furthermore, the method has been successfully applied to genomic selection. On the basis of supervised dimension

reduction strategy, the developed methods can handle all polymorphisms across the entire genome simultaneously. We have demonstrated the superior performance of the methods with computer-based simulation studies, and the utility of the methods using real data analysis.

W426: Genomic Selection and Genome-Wide Association Studies

Haplotype-based Genomic Prediction of Breeds Not in Training

Xiaochen Sun[1], Hailin Su[1], Prashanth Boddhireddy[2] and Dorian J. Garrick[1], (1)Department of Animal Science, Iowa State University, Ames, IA, (2)Zoetis, Kalamazoo, MI

In this study we characterized breed differences and compared genomic prediction among multi-breed beef cattle populations using haplotype alleles constructed from phased genome-wide SNP genotypes. The Bovine 50K SNP genotypes of 1,872 Angus, 1,341 Charolais, 1,229 Gelbvieh, 3,221 Hereford and 3,252 Simmental animals were phased and imputed within-breed to the density of the 700K Bovine HD panel. Haplotype alleles were constructed within consecutive non-overlapping genomic windows of 1 million-bp (Mbp) for real 50K and imputed HD genotypes, and 100 kilo-bp (Kbp) for imputed HD genotypes. Alleles with a frequency within breed >1% were defined as common. The 5 breeds shared < 1% of 1Mbp common alleles, but > 9% of 100Kbp common alleles. The first three principal components of haplotypes characterized breed differences better than SNP genotypes. Genomic prediction was performed for birth weight, weaning weight, yearling weight, back fat thickness and ribeye area using deregressed proofs with parental averages added back. Method BayesC was used to estimate SNP effects (SNP model) or effects of common haplotype alleles (haplotype model). Prediction accuracies of the 100Kbp haplotype model were similar or slightly higher than the SNP model when training populations included animals from the same breed as prediction candidates, but were significantly higher for certain traits and breeds when training populations excluded the breed of the prediction candidates. These results were supported by the observation that linkage disequilibrium (LD) was much higher or complete between multi-allelic haplotypes and bi-allelic SNPs across breeds, compared to pair-wise LD between single SNPs. Furthermore, the haplotype model was computationally more efficient than the SNP model because the number of common 100Kbp haplotype alleles was less than half the number of HD SNPs. In conclusion, the haplotype model is a more accurate and efficient alternative to the SNP model for genomic prediction when animals of the same breed as prediction candidates are not available for training.

W427: Genomic Selection and Genome-Wide Association Studies

Genome-Wide Association Provides Insights into Exploitable Natural Variation for Developing Climate Smart Chickpeas

Rajeev K Varshney, ICRISAT, Hyderabad, India

Chickpea (*Cicer arietinum* L.) is an important food legume crop for global food security in developing countries, but productivity is severely limited due to biotic and abiotic stress. In order to exploit the genome sequence for chickpea improvement, 429 chickpea genotypes of diverse origin were re-sequenced at 5X to 13X coverage. Using this data, we identified genome-wide variations including 4.9 million single nucleotide polymorphisms (SNPs), 596,100 Indels, 4,931 copy number variations, 60,742 presence absence variations and 70,159 structural variations. In addition, this analysis provides insights in to population structure, genetic diversity, gene loss, domestication and selection sweeps. For identifying marker-trait associations (MTAs) that can be applied in chickpea improvement, we used genome-side SNP data coupled with phenotyping data generated for drought and heat tolerance related traits from 1–6 seasons and 1–3 locations in India and three locations in Africa. In total, 333 MTAs were identified for 13 traits, of which 249 were for drought tolerance related traits, 67 for heat tolerance related traits and 17 for heat tolerance index. Based on the physical position of associated SNP loci on the CDC Frontier reference genome, candidate genes like TIME FOR COFFE, RELATIVE OF EARLY FLOWERING 6, disease resistance etc., were identified. We report haplotype patterns shaping the genomic background of elite varieties used by farmers, and marker trait associations for drought and heat tolerance that can be deployed to accelerate chickpea improvement.

W428: Genomic Selection and Genome-Wide Association Studies

Combining GWAS and Genomic Prediction using Sequence Information on 5000 Holstein Bulls, effect on Prediction Accuracy

Roel F. Veerkamp, Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Wageningen, Netherlands and Aniek Bouwman, ABGC, Wageningen Livestock Research, Wageningen, Netherlands

Our earlier works (van Binsbergen et al, 2015) showed that using whole genome sequence information did not improve the accuracy for genomic prediction within the Holstein population. In this study we therefore compared genomic prediction using selected SNP from single SNP GWAS with imputed whole genome information (from run4 of the 1000 bull genomes project) to genomic prediction obtained using either the full sequence, 50k or BovineHD SNP chips. The GWAS used highly accurate deregressed proofs from 3416 training bulls, all progeny tested bulls for protein yield, interval first to last insemination and SCC . The analyses were performed in the GCTA package. The genomic relationship matrix was included to account for population structure and was based on the SNP of the bovine HD chip. After the GWAS, the relevance of different selected SNP sets were tested by estimating the h^2 and the prediction accuracy in 2287 validation animals, using the GRM calculated with different SNP subsets. From the 30 million SNP in the sequence information, only 13,789,029 where segregating in the Holstein population. For protein yield 2,194 SNPs were significant ($-\log_{10}(p) > 5$), and 28 (160) of those were present on the 50k (HD) SNP chips. Within the Holstein population prediction accuracy as well as the h^2 were lower when SNPs were selected based on the GWAS, and there was no advantage in comparison with the 50K or BovineHD SNPchips. But more selective selection procedures and training population might be required to benefit from the precision of full sequence genotypes.

W429: Genomics of Genebanks

Are You Getting What You Ordered from Your Genebank? Fingerprinting of the Clonal Potato and Sweetpotato Collections at the International Potato Center

David Ellis[1], Noelle Barkley[2], Rene Gomez[3], Genoveva Rossel[1], Julian Soto[1], Ronald Robles[1], Joseph Coombs[4], Andrzej Kilian[5] and David Douches[6], (1)International Potato Center, Lima, Peru, (2)1International Potato Center,, Lima, Peru, (3)International

Potato Center, Av. La Molina 1895, Lima, Peru, Lima, Peru, (4)Michigan State University, East Lansing, MI, (5)Diversity Arrays Technology Pty Ltd, Canberra, Australia, (6)University of Canberra, Bruce, Australia
Quality assessments using morphological descriptors of germplasm collections are usually not done except during regeneration for seed accessions or during field grow-out for clonal crops. Genetic analyses are rarely available, even for clonal crops maintained as vegetatively propagated genetically stable unique genotypes. The International Potato Center (CIP) maintains one of the largest in vitro clonal collections with over 11,000 accessions. Accessions are also maintained as field plantings from original material providing an exceptional opportunity to assess if the tissue culture accessions are the same genotype after 30+ years. The SolCAP SNP Array was used to fingerprints 4,350 accessions in the active potato collection while simple sequence repeats (SSRs), followed by DArTseq was used for the over 5,000 accessions in the active sweetpotato collection. The original mother plants and original descriptor data was used for comparison to confirm in vitro material was true-to-type. These studies found approximately 85% of the accessions in the in vitro collections were true-to-type and the original material generally matched primary descriptors from the original collections. Although our working hypothesis is that mistakes were made early in the isolation of accessions in vitro, the work clearly highlights the need for such studies. The generation of unique fingerprints has already yielded unanticipated results by highlighting errors in determining the ploidy levels of the potato germplasm (2X, 3X, 4X, and 5X). Long-term, genetic fingerprints will be a critical component of the CIP quality management system to ensure genetic integrity of genebank accessions now and into the future.

W430: Genomics of Genebanks

Genomic Characterization of Domestication and Genome-Wide LD in Beet

Brigitte Mangin[1], Karine Bounan Henry[2], Brigitte Devaux[2], Glenda Willems[3], Pierre J. Devaux[2] and **Ellen Goudemand**[2], (1)Laboratoire des Interactions Plantes Micro-organismes (LIPM) - INRA/CNRS, Castanet-Tolosan, France, (2)Florimond Desprez, Cappelle-en-Pévèle, France, (3)SESVANDERHAVE N.V., Tienen, Belgium
The genetic diversity of a worldwide beet population containing 2035 accessions was analyzed using four different approaches. Three of the methods gave a very coherent picture of the population structure. Fodder beet and sugar beet accessions were grouped together, separated from garden beets and sea beets, reflecting well the origins of beet domestication. Then, linkage disequilibrium (LD) of this worldwide population was compared to the LD of one set of 1338 elite sugar beet lines. The usual measure (r^2) was used, and compared with others that correct for population structure and relatedness (r_S^2, r_V^2, r_{VS}^2). The LD as measured by r^2 persisted beyond 10 cM within the elite panel and fell below 0.1 after less than 2 cM in the worldwide population, for almost all chromosomes. With correction for relatedness, LD decreased under 0.1 by 1 cM for almost all chromosomes in both populations, except for chromosomes 3 and 9 within the elite panel. In these regions, the larger extent of LD could be explained by strong selection pressure.

W431: Genomics of Genebanks

Bean Adapt: The Genomics of Adaptation during Crop Expansion in Bean

Roberto Papa, Marche Polytechnic University, Ancona, Italy, Scott A. Jackson, University of Georgia, Athens, GA, Paul Gepts, University of California, Davis, CA, Andreas Graner, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and Alisdair R. Fernie, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
BEAN_ADAPT is a three year project funded through the 2nd ERA-CAPS call, ERA-NET for Coordinating Action in Plant Sciences.
The main aim of this project is to dissect out the genetic basis and phenotypic consequences of the adaptation to new environments of the common bean (*Phaseolus vulgaris* L.) and its sister species, the runner bean (*Phaseolus coccineus* L.), through the study of their introduction, from their respective centers of domestication in the Americas, and expansion through Europe, as a recent and historically well-defined event of rapid adaptation.
A large collection (11,500 accessions of both species) from three major genebanks, will be characterised by genotyping-by-sequencing (GBS), to define the population structure and to obtain subsets of genotypes for phenotyping (field and growth chamber) and for a deeper genomic–transcriptomic–metabolomic characterisation. We will use a multidisciplinary approach: genomics (WGS and RNAseq), population/ quantitative genetics, biochemistry, plant physiology on the subset of samples. Differential expression analysis, analysis of the co-expression patterns, and GWAS will be used to identify genes and metabolites putatively associated with adaptation, while genotypic information obtained from RNA-seq data will be used, with GBS and WGS data, to test for signatures of selection.
Among the main outcomes of the project are the development in *P. vulgaris* of haplotypes of all 10,000 accessions (HapBean), along with associated information and seed stocks, which will represent a unique tool for plant scientists and breeders. For *P. coccineus*, we will also have a well-defined set of information that will constitute the foundation for the development and application of its genomic resources.

W432: Genomics of Genebanks

Developing Climate Resilient Wheat: Manipulating Major Genes and Exploiting Novel Diversity

Alison R Bentley, The John Bingham Laboratory, NIAB, Cambridge, United Kingdom
The NIAB wheat pre-breeding program exists as a “proof-of-concept” for translating plant science research into pre-commercial breeding material for UK wheat improvement. As part of this program we are investigating key genes controlling flowering time, a trait underpinning adaptation to the environment. In breeding, flowering time can be manipulated via major gene combinations to tailor wheat varieties to target environments. Flowering time work at NIAB encompasses the development of precise genetic stocks (e.g. Bentley et al., 2013 JXB 64:1783-1793) to understand broad environmental response as well as to reveal developmental responses and tradeoffs associated with temperature and drought stress. Next generation mapping resources, including the NIAB Elite Multi-parent Advanced Generation InterCross (MAGIC) population (Mackay et al., 2014 G3 4:1603-1610) also elucidate genetic loci involved in fine-tuning environmental adaptation.
In addition to the action of major genes for environmental adaptation, we are exploiting genetic diversity from wheat’s wild relatives. As part of the UK public-sector wheat pre-breeding programme WISP (http://www.wheatisp.org), we are mining *ex situ* collections of wheat’s D-genome donor, *Aegilops tauschii*. New diversity from 50 independent *Ae. tauschii* accessions from a wide genetic and ecogeographical range has been captured via targeted re-synthesis of the hexaploid (AABBDD) wheat genome. This diversity is exploitable through a structured pre-breeding pipeline which aims to produce 10,000 derived lines in elite backgrounds. High density genotyping is used in concert with targeted phenotypic

assessment and a number of genotyping resources (e.g. Winfield et al. 2015 PBJ in press) have been, and continue to be, developed in support of this. A new component of the programme is the incorporation of favourable abiotic stress traits from a small collection of Hourani tetraploid wheat landraces from Syria.

W433: Genomics of Genebanks

Expanding the Gene Pool for Crop Improvement Using the Genomes of Crop Wild Relatives

Robert Henry, University of Queensland QAAFI, Brisbane, Australia

Sequencing the genomes of crop wild relatives provides a tool for characterizing the germplasm in the wild and in genebanks. This approach supports improved conservation and management of plant genetic resources *in situ* and *ex situ*. Sequencing of wild rice populations has revealed evolutionary relationships and will useful to rice breeders in accessing more diverse sources of useful genes. Key populations have been identified as requiring conservation in the wild and further collection to complement the material held in gene banks. Extensive transcriptome sequencing in wheat seeds of diverse germplasm has shown important diversity that explains variation in key functional traits. This analysis identifies the genes that have been the basis of human selection in domestication and breeding and guides more effective and targeted breeding in the future. Sequencing the genomes of wild relatives of long lived species such as trees has the potential to greatly accelerate the breeding of these species.

W434: Genomics-Assisted Breeding

Welcome & Introduction

Rajeev K Varshney, ICRISAT, Hyderabad, India

W435: Genomics-Assisted Breeding

Using Molecular Annotations to Find the SNPs for Breeding

Edward S. Buckler[1,2], Eli Rodgers-Melnick[2] and and the Maize Diversity Project[2], (1)USDA-ARS, Ithaca, NY, (2)Institute for Genomic Diversity, Cornell University, Ithaca, NY

Over much of the last two decades, tremendous strides have been made both in molecular genetics and quantitative genetics. However, during most of this time quantitative genetics mostly used molecular genetics as nearly inexhaustible supply of molecular markers. Detailed examinations of chromatin structure and genome conservation are providing insights into the thousands of variation that contribute to phenotypic variation, while allowing the millions neutral variants to be ignored. We will show how a detailed examination of 10's millions of variants in maize is providing insights in heterosis, adaptation, and the gene classes responsible for adaptation.

W436: Genomics-Assisted Breeding

Genome-Based Establishment of a High-Yielding Heterotic Pattern for Hybrid Wheat Breeding

Jochen Christoph Reif, IPK - Gatersleben, GATERSLEBEN, Germany

Hybrid breeding is a promising approach in selfing species to boost yield and yield stability. The success of hybrid breeding depends crucially on the clustering of germplasm into heterotic groups, and on the identification of a high-yielding heterotic pattern. Here we have developed a three-step approach to identify a promising heterotic pattern which comprises following elements: (1) The full hybrid performance matrix is compiled using genomic prediction. (2) A high-yielding heterotic pattern is searched based upon a developed simulated annealing algorithm. (3) The long-term success of the identified heterotic pattern is assessed. The three-step approach was successfully implemented and evaluated using a phenotypic and genomic wheat data set comprising 1,604 hybrids and their 135 parents. We show that hybrid wheat breeding based on the identified heterotic pattern boosts grain yield through the exploitation of heterosis and enhances recurrent selection gain. Thus, the framework for the recognition of heterotic groups developed in our study represents a central step forward to initiate hybrid breeding programs with the final goal of meeting the global challenges of an increasing demand for food, feed, and fuel.

W437: Genomics-Assisted Breeding

Genomics and Drought Tolerance in Wheat

Peter Langridge, Australian Centre of Plant Functional Genomics, Adelaide, Australia

Water deficit or drought stress is a major limitation to crop production globally, particularly for wheat where average global yields of rain-feed production systems are below 1.5 tonnes per hectare. Wheat breeders have used a wide range of technologies to successfully breed varieties that perform well under the growth conditions for their target environments but they are always seeking new opportunities to enhance rates of genetic gain. Under drought, yield is determined by the integration of variable levels of water deficit across the developmental life of the crop. Genomics technologies were seen as a path to understanding the genetic and environmental complexity of drought stress. However, to be relevant to breeding programs, genomic studies must consider the nature of drought stress in the target environment and use plant material and phenotyping techniques that relate to field conditions.

W438: Genomics-Assisted Breeding

Genomic Prediction in Wheat and Maize Breeding Populations

Jose Crossa[1], Paulino Perez-Rodriguez[2], Gustavo de los Campos[3], Jessica Rutkoski[4], Jesse Poland[5], Ravi P. Singh[1], Jaime Cuevas[6], Xuecai Zhang[1] and Yoseph Beyene Aydagn[7], (1)CIMMYT, Mexico DF, Mexico, (2)Colegio de Postgraduado, Texcoco, Mexico, (3)Michigan State University, East Lansing, MI, (4)Cornell University & CIMMYT, Ithaca, NY, (5)Kansas State University, Manhattan, KS, (6)Universidad de Quintana Ro, Chetumal, Mexico, (7)CIMMYT, Nairobi, Kenya

Selection of the bandwidth parameter *h* in the Gaussian kernel is a challenging problem. Recently a Bayesian method for selecting the bandwidth parameter (*h)* of a Gaussian kernel was developed by assigning a prior *p*(*h*) and obtaining a posterior point estimate of *h*. Also recently, a marker x environment interaction model (MxE) in the context of the GBLUP model was proposed that allows decomposing marker effects and genomic

variance into components that are stable across environments (main effect) and components that are environment specific (interaction terms). In the MxE model the genetic covariance between any pair of environments can be represented by the variance of the main effect; therefore it is restricted to being positive (homogeneous or heterogeneous across environments). The approach for selecting the bandwidth parameter *h* can be implemented in conjunction with the MxE model to examine accuracies of different prediction problems. Furthermore, the reaction norm model, also recently developed for assessing genotype x environment interaction (GxE), allows genomic prediction using highly dimensional marker, pedigree and environmental covariables matrices. Results of two wheat data sets using the MxE model with the Gaussian kernel Bayesian estimation of *h*, show increases in prediction accuracies over the standard MxE of about 7% and10%. Other results include the analyses of GxE incorporated in the reaction norm model using 20,000 wheat lines genotyped with genotyping-by sequencing. For predicting about 500 wheat lines in several international environments (in two different years), environmental covariables of the environments to be predicted were included. The prediction accuracy of these international environments shows promising results.

W439: Genomics-Assisted Breeding

Genomics-Assisted Breeding for Nematode Resistance in Soybeans

Ki-Seung Kim[1], Dan Qiu[1], Li Song[1], Tri D. Vuong[1], Juexin Wang[1], Trupti Joshi[1], J. Grover Shannon[2] and Henry T. Nguyen[1], (1)University of Missouri, Columbia, MO, (2)University of Missouri, Portageville, MO

Soybean nematodes, including soybean cyst nematode (SCN), root-knot nematode (RKN), and reniform nematode (RN) are the most destructive pests in soybean production. Over 100 SCN resistant germplasm and several resistant QTL are known. However, more than 90% of commercialized SCN-resistant soybean cultivars in the USA carry only *rhg1* from PI 88788. In the past few years, over 1,000 Plant Introductions (PIs) (MG 000-V) were evaluated for nematode resistance. Several PIs with broad-based resistance to multiple SCN HG types and resistance to RKN and/or RN were identified. New SCN resistant QTL in PI 437654 and PI 567516C have been discovered. Whole genome sequencing and haplotype analysis are also being employed to dissect genomic locations and QTL/genes associated with resistance to the multiple-nematodes. To facilitate marker-assisted breeding (MAB), two gene specific SNP markers were developed for *rhg1* which can differentiate eight copies (as in PI 88788), three copies (as in Peking), or one copy (as in Williams 82). Other SNP markers were developed which distinguish presence and absence of *Rhg4* from Peking. Fine-mapping, cloning, and functional marker development for the QTL in PI 437654 and PI 567516C are underway. A major RKN resistant QTL in PI 438489B was mapped within a 27.6 kb region by whole genome resequencing and SNP markers were developed from two candidate genes. The overall goal of our program is to discover new sources, identify novel QTL/genes providing resistance to multiple nematode species, and develop functional markers as well as new varieties with broad spectrum nematode resistance through MAB.

W440: Genomics-Assisted Breeding

The Genomic & Open-source Breeding Informatics Initiative

Kelly Robbins[1], Edward S. Buckler[2], Jean-Luc Jannink[1], Tobias Kretzschmar[3], Lukas Mueller[4], Yaw A. Nti-Addae[1], Michael S. Olsen[5], Mark E Sorrells[1], Qi Sun[6], Rajeev K Varshney[7] and Susan McCouch[1], (1)Cornell University, Ithaca, NY, (2)USDA-ARS-Cornell University, Ithaca, NY, (3)International Rice Research Institute, Los Baños, Philippines, (4)Boyce Thompson Institute for Plant Research, Ithaca, NY, (5)CIMMYT, Nairobi, Kenya, (6)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (7)ICRISAT, Hyderabad, India

In the last ten years, genotyping costs have dropped significantly, making feasible powerful new breeding approaches that can take advantage of the vast amounts of genomic data that have been generated in staple crops such as rice, wheat, maize, sorghum, and chickpea. The Genomic & Open-source Breeding Informatics Initiative (GOBII) is the first large-scale public-sector effort to enable systematic application of high-density genotypic information to the breeding of staple crops in the developing world. The project will develop and implement genomic data management systems to enhance the capacity of public-sector breeding programs to deliver increased rates of genetic gain in South Asia and Sub-Saharan Africa. The genomic data management systems will include databases, analysis pipelines, and decision support tools for plant breeders.

W441: Genomics-Assisted Breeding

Summary and Wrap-up

Rajeev K Varshney, ICRISAT, Hyderabad, India

W442: Genomics of Non-Classical Model Animals

Genome Rearrangements Contribute to Vocal Learning in Birds: A Model to Study Evolution of Speech?

Denis M. Larkin[1,2], Marta Farré-Belmonte[2], Joana Damas[1] and Darren K. Griffin[3], (1)Department of Comparative Biomedical Sciences, Royal Veterinary College, University of London, London, United Kingdom, (2)Royal Veterinary College, University of London, London, United Kingdom, (3)School of Biosciences, University of Kent, Canterbury, United Kingdom

Homologous synteny blocks (HSBs) and evolutionary breakpoint regions (EBRs) in mammalian chromosomes are enriched for distinct DNA features, contributing to distinct phenotypes. To reveal HSB and EBR roles in avian evolution, we performed a sequence-based comparison of 21 avian and five outgroup species using recently sequenced genomes across the avian family tree. We identified EBRs and HSBs in ancestral bird, archosaurian (bird, crocodile, dinosaur), and reptile chromosomes. Archosaurian-specific HSBs were enriched for genes that function in retina structures, while avian-specific HSBs for genes involved in limb development. The average rate of rearrangements in birds is ~1.25 EBRs per million years; however, bursts of genomic reorganization occurred in several avian lineages. For example, the origin of Neognathae was accompanied by an elevated rate of chromosome rearrangements. Intriguingly, all vocal-learning species had significantly higher rates of rearrangements than those of close vocal nonlearning relatives and even higher relative to all vocal nonlearning species. EBRs leading to budgerigar after the divergence from the Passerimorphae ancestor tended to reshuffle genes involved in forebrain development. Remarkably, the forebrain development GO term was enriched in avian and archousaurian msHSBs as well. However, the gene pathways affected by EBRs and msHSBs were different. This example indicates that both EBRs and msHSBs are sources of variation used by natural selection to form complex phenotypes over evolutionary time. Our findings provide novel evolutionary insights into genome evolution in birds, particularly how

chromosome rearrangements contributed to formation of phenotypes that make vocal-learning birds excellent models to study evolution of speech.

W443: Genomics of Non-Classical Model Animals

Genetic Contributions from Archaic Humans and their Effects on Human Fitness

Emilia Huerta-Sanchez, Molecular Cell Biology, School of Natural Sciences, University of California, Merced, Merced, CA

Comparisons of DNA from archaic and modern humans show that modern and archaic humans interbred, and in some cases received an evolutionary advantage from doing so. This process, adaptive introgression (AI), may lead to a faster rate of adaptation than is predicted from models with mutation and selection alone. Within the last couple of years, a series of studies have identified regions of the genome that are likely examples of AI. One clear example of AI involves the EPAS1 (HIF-2α) gene that had the most striking signature of positive selection in Tibetans, and the selected haplotype at this selected locus was most likely derived from an archaic human population (Denisovan) that had interbred with the ancestral Tibetan population. Other examples of AI are found in genes involved in immune function, metabolism and skin and hair pigmentation, suggesting that the selective agents are different pathogens, diets and temperatures, respectively. Here, we examine known examples of AI to discover new statistics that can specifically differentiate between introgression and adaptive introgression, and we use simulations to investigate the false positive rate of these statistics under null models that include introgression in the absence of selection

W444: Genomics of Non-Classical Model Animals

The Role of Chromosomal Rearrangements in Speciation and Adaptation in Primates– New Tools into Human Welfare Studies?

Marta Farré-Belmonte, Royal Veterinary College, University of London, London, United Kingdom, Anna Ullastres, Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain, Laia Capilla, Institut de Biomedicina i Biotecnologia (IBB), Universitat Autònoma de Barcelona, Bellaterra, Spain and Aurora Ruiz-Herrera, Institut de Biotecnologia i Biomedicina (IBB), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Comparative genomics has become a powerful tool to study the links between genome evolution and phenotypes. Using genomes of several primate species (human, chimpanzee and macaque) we detected regions of conserved synteny (HSBs) surrounded by evolutionary breakpoint regions (EBRs). Chromosomal rearrangements might act as barriers to gene flow due to a suppression of recombination within reorganized regions, thus facilitating the accumulation of genetic incompatibilities that eventually could lead to speciation. Favouring this view, in human and chimpanzee, we found lower recombination rates in inverted chromosome regions than in collinear chromosomes. Interestingly, human chromosomes 1 and 18, containing human-specific inversions, have genes differentially expressed in cerebral cortex (e.g., USP14 and HSRTS-BETA in human chromosome 18). When we extended these analyses to the macaque genome by estimating recombination rates we found that macaque chromosomes 5 and 9, containing macaque-specific inversions, had lower recombination rates than the rest of the chromosomes. Further, analysing the gene content, we found that defence-response genes, differentially expressed between human and other primates in the reproductive tract, were enriched in rhesus macaque EBRs. Moreover, a cluster of alpha-defensins, which are positively selected in the macaque lineage, was found in the inverted region of macaque chromosome 5. In summary, the study of chromosome evolution in primates is revealing as a powerful tool in pointing to the specific genes and evolutionary mechanisms that could be related to phenotypic differences between humans and their close primate relatives.

W445: Genomics of Plant Development

Parasitic Plants Signal Network Analysis

Claude dePamphilis, Penn State University, University Park, PA

W446: Genomics of Plant Development

A Maize Kernel Mutant that Lacks a Putative Function in Cell Division

Nelson Garcia, Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ and **Joachim Messing**, Rutgers University, Piscataway, NJ

We have identified a new defective kernel (dek) mutant in maize called dek34-Dsg1 with apparent abnormalities in cell division and differentiation. The mutation is caused by an insertion of a GFP-tagged Ds transposable element (DsG) in a gene that is predicted to encode Tel2 Interacting Protein 2 (TTI2). TTI2 is a member of the triple-T (TTT) complex that regulates the maturation and stability of the phosphatidylinositol 3-kinase-related protein kinase (PIKK) family. Early cell divisions during embryo development seem to proceed normally, but eventually arrests at the transition/dermatogen stage. The endosperm is also severely underdeveloped, has highly reduced protein, and lacks histologically distinct compartments. These phenotypes are consistent with mutation in the PIKK member Target of Rapamycin (TOR), which is a major regulator of cell growth and division in response to nutrients. Furthermore, segregation analyses from selfed and reciprocal crosses did not indicate any gametophytic maternal effect, but showed reduced pollen transmission which is consistent with a previously characterized Ataxia Telangiectasia Mutated (ATM) mutant in Arabidopsis, another PIKK member that is involved in DNA damage repair. Taken together, these results show strong evidence for the role of dek34-Dsg1 in PIKK function.

W447: Genomics of Plant Development

Gene Networks in Plant Biology: Approaches in Reconstruction and Analysis

Yupeng Li, Stephanie A. Pearl and Scott A. Jackson, University of Georgia, Athens, GA

Even though vast amounts of genome-wide gene expression data have become available in plants, it remains a challenge to effectively mine this information for the discovery of genes and gene networks, for instance those that control agronomically important traits. These networks reflect potential interactions among genes and, therefore, can lead to a systematic understanding of the molecular mechanisms underlying targeted biological processes. We discuss methods to analyze gene networks using gene expression data, specifically focusing on four common statistical

approaches used to reconstruct networks: correlation, feature selection in supervised learning, probabilistic graphical model, and meta-prediction. In addition, we discuss the effective use of these methods for acquiring an in-depth understanding of biological systems in plants.
http://dx.doi.org/10.1016/j.tplants.2015.06.013

W448: Genomics of Plant Development

Rose Genomics: Insights into Flower Development and Function

Mohammed Bendahmane[1], J. Just[1], O. Raymond[1], L. Francois[1], Xiaopeng Fu[1,2], P. Vergne[1], D. Annick[1], A. Roccia[1], S-H. Yang[1], S. Baudino[3], JC. Caissard[3], Jl. Magnard[3], J. Gouzy[4], S. Carrere[4], L. Legrand[4], M. Le Bris[5], B. Goveto[5], J. Salse[6], A. Boualeme[7], M. Benhamed[7,8] and A. Bendahmane[7], (1)Ecole Normale Supérieure, Lyon, France, (2)Huazhong Agricultural University, Wuhan, China, (3)Université Jean Monnet, Saint-Etienne, France, (4)LIPM-INRA Toulouse, Castanet-Tolosan, France, (5)Aix-Marseille University, Marseille, France, (6)INRA Clermont-Crouel, Clermont-Ferrand, France, (7)IPS2-Paris-Sud University, Orsay, France, (8)King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Rose is one of the most cherished domesticated plants in human history. During centuries, generations of rose breeders had fastidiously selected the showy and desirable traits of *Rosa* species based on keen and meticulous observation. Several traits, involving mainly floral quality (ie. recurrent flowering, flower form and double flowers, scent, …), are of high economic importance. The molecular and genetic mechanisms controlling these characters remain poorly understood. Some of these important characters that are difficult to address in other model species, such as *Arabidopsis*, can be studied in the rose. Besides its economic importance, the rose is well suited to be a model organism for woody ornamental species. Moreover, the rose has a relatively small genome size (about 560 Mbp) and it has a short life cycle for a perennial woody plant. During the past years, we and our collaborators have generated a number of molecular, genomic and biotechnology tools such as an efficient and reproducible genetic transformation as well as a database and a WEB interface that provide useful information on *Rosa sp.* expressed genes, with thorough annotation and an overview of expression patterns for transcripts with good accuracy; the latest represented a valuable prerequisite to the sequencing of the rose genome, currently in progress. We are using a multi-approach scale to help understanding the molecular and genetic mechanisms associated with some of these traits and more specifically flower initiation and development as well as scent biosynthesis. Recent advances will be presented and discussed

W449: Genomics of Plant Development

Genetic Diversity within the Vitamin E Biosynthetic Pathway of Sunflower

Linchay Janine Daniels, Agricultural research council, Biotechnology Platform, Pretoria, South Africa

Sunflower (*Helianthus annuus*) is one of the world's most important oil seed crops. It's oil's oxidative stability and shelf life is conferred, amongst other, by it's vitamin E (tocopherol) content. The biosynthetic pathway of tocopherol however, has yet to be fully characterized in sunflower. This research aims to: (1) identify the tocopherol biosynthetic gene homologs in public bioinformatic databases in order to construct a biosynthetic pathway and (2) to identify and characterize the genetic diversity of tocopherol genes using next generation DNA sequencing technologies. The identified candidate genes were assigned to Gene Ontology (GO) terms, mapped in TAIR and remapped in KEGG Orthology (KO) database. The resultant reference pathway, "*Ubiquinone and other Terpenoid Quinone biosynthesis pathway*" was further analyzed and the tocopherol pathway was constructed. This pathway was used to select the genes TC, TMT, HPT and HPPD responsible for the production of tocopherol derivatives. The four specific gene sequences were retrieved from NCBI and aligned to the *Helianthus annuus* draft genome to determine consensus sequences for each gene. Sequencing revealed that 74,5M reads mapped for the gene TC and 64,3M for TMT to *Helianthus annuus* on NCBI with a 100% and 96% nucleotide similarity, respectively. The gene sequences HPPD (1900) and HPT (4,7M) mapped against *Lactuva sativa*, a close relative of sunflower. The gene HPPD showed 80% nucleotide similarity while HPT had 91% similarity. These genes will be further analyzed for 104 sunflower accessions to develop SNP markers for oxidative stability and shelf life for future breeding.

W450: Genomics of Tissue Regeneration in Plants and Animals

Stem Cells and Regeneration in Planarians

Ricardo M Zayas, San Diego State University, San Diego, CA

Regeneration of missing body parts has long fascinated biologists, yet the mechanisms underlying regenerative processes remain poorly understood. My laboratory uses the freshwater planarian, a classic model of regeneration studies, as a model to examine the molecular mechanisms underlying regeneration. Planarians are able to completely regenerate entire worms and form lost body parts from very small body pieces. These organisms are endowed with a population of adult pluripotent stem cells that support their capacity for regeneration, which provides an excellent opportunity to identify genes involved in stem cell maintenance, proliferation, and differentiation. In my talk, I will briefly introduce planarians and discuss our work aimed at understanding how regeneration of specific organs is achieved in these fascinating organisms.

W451: Genomics of Tissue Regeneration in Plants and Animals

Genetic Control of Distal Stem Cell Fate within Root and Embryonic Meristems

Brian Crawford, Section of Cell and Developmental Biology, La Jolla, CA

The root meristem consists of populations of distal and proximal stem cells, and an organizing center known as the quiescent center. During embryogenesis, initiation of the root meristem occurs when an asymmetric cell division of the hypophysis forms the distal stem cells and quiescent center. We have identified *NO TRANSMITTING TRACT (NTT)* and two closely related paralogs as being required for the initiation of the root meristem. All three genes are expressed in the hypophysis and their expression is dependent on the auxin-signaling pathway. Expression of these genes is necessary for distal stem cell fate within the root meristem, while misexpression is sufficient to transform other stem cell populations to a distal stem cell fate in both the embryo and mature roots.

W452: Genomics of Tissue Regeneration in Plants and Animals

Axon Regeneration in *C. elegans* : Genetics and Genomics

Andrew Chisholm, UCSD, San Diego, CA

Regenerative regrowth of axons after damage is widespread throughout the animal kingdom. In adult mammals, extensive axon regeneration is possible in the peripheral nervous system, whereas the central nervous system has little or no regenerative capacity. Simple models of axon regeneration facilitate gene discovery and mechanistic studies of novel regulators of axon regeneration. The nematode *C. elegans* has highly stereotyped nervous system composed of 302 neurons that exhibits robust regeneration after laser axotomy. Key injury-induced signaling pathways such as the DLK-1 MAP kinase cascade play conserved roles in axon regeneration in *C. elegans* and in mammals. Large-scale screens have identified numerous regulators of gene expression as required for axon regeneration, including the CELF (CUGBP and Elav-like) family RNA-binding protein UNC-75 (Chen et al., 2011, Neuron 71: 1043). Genomic and transcriptomic analysis of UNC-75 targets in axon regeneration has revealed a conserved regulatory pathway required for efficient axon extension.

W453: GMOD
Introduction to GMOD
Scott Cain, Ontario Institute for Cancer Research, Medina, OH

W454: GMOD
Apollo: Newest Features in Collaborative Genome Curation
Monica C. Munoz-Torres[1], Nathan A Dunn[1], Deepak Unni[2], Colin Diesh[2], Christine G. Elsik[2], Ian Holmes[3] 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
Apollo enables collaborative, real-time curation (akin to Google Docs) of genomic elements using both structural and experimental information. Built on top of the JBrowse framework, Apollo is composed of a web-based client, an annotation-editing engine, and a server-side data service. Users can visualize gene models, protein alignments, and expression and variant data to conduct structural and/or functional annotations. In our most recent release, version 2.0.x, the improved architecture allows users to more easily query data and build extensions, supports multiple organisms per server, and allows additional types of sequence annotations based on the Sequence Ontology. The new, removable side-dock offers detailed view of annotations, sequences, and organisms, a new reporting structure, and WebSocket support to improve real-time communication. The new Grails framework (Spring / Hibernate / Groovy) in the server more robustly scales a single server over multiple organisms while better supporting additional curators. Apollo's entire secure REST API is exposed, allowing genomic features to be injected into Apollo from an automated curation process or organization-specific metadata to be extracted directly from Apollo using a SQL query or REST. The new version offers improved features, including the ability to bring together 2 or more scaffolds in order to annotate genes split across them, and increases the ability to customize and integrate Apollo into modern curation pipelines. During this demonstration we will introduce the new architecture, highlight advantages for users, and detail our future plans. **Project Website**: http://genomearchitect.org/ **Source Code**: https://github.com/GMOD/Apollo **License**: Berkeley Software Distribution (BSD) License at https://github.com/GMOD/Apollo/blob/master/LICENSE.md

W455: GMOD
Development of an Open, Community Driven, Central Database for Microbial Genetic Knowledge in Wikidata
Tim Putman, Scripps Research Institute, La Jolla, CA

W456: GMOD
Galaxy Community Update
Dave Clements, Johns Hopkins University, Eugene, OR
Galaxy is a widely used and deployed data integration and analysis platform for life science research. This talk will briefly introduce the platform and then discuss usage, deployment options, and features with an emphasis on what's new in the past year. Recent efforts have focused on user interface enhancements, dataset management, easier deployment and tool management, cloud support. and integrating Galaxy analysis with ad hoc tools such as R and other scripting tools.
For an in-depth demonstration of Galaxy, please attend the *Galaxy for SNP and Variant Data Analysis* workshop on Tuesday at 4pm in the California Room.

W457: GMOD
GMOD and Chado: Present and Future
Scott Cain, Ontario Institute for Cancer Research, Medina, OH

W458: GMOD
GMOD Community Roundtable
Scott Cain, Ontario Institute for Cancer Research, Medina, OH

W459: Graft Genetics and Genomics
Interfamilier Grafting Using a Plant Genus *Nicotiana*
Michitaka Notaguchi, JST ERATO Higashiyama Live-Holonics Project, Nagoya, Japan
Plant grafting has been an important technique in horticulture as well as in agriculture to perpetuate clones, to obtain the benefits of certain rootstocks such as disease resistance and tolerance of unfavorable soil conditions and to control fruit production. However, like as immunological resistance found in animals, plant grafting also has a limitation in combinations to graft. In general, it has been thought that grafting can be done successfully between the same species, genus, and family, but not between different families because of graft incompatibility. Recently, we found that a *Solanaceae* species, *Nicotiana*, can be grafted onto a *Brassicaceae*, Arabidopsis, where a population

of mRNA was transported from Arabidopsis stocks to *Nicotiana* scions (Notaguchi et al., 2015). We further showed that *Nicotiana* has a capability to reduce or overcome the graft incompatibility. *Nicotiana* was successfully grafted with a wide range of plants, including magnoliids, monocots and dicots. In our anatomical analyses, vasculature reconstruction at the graft junction, which is normally observed in grafting of compatible combinations, was not observed. But, alternatively, parenchyma tissues were proliferated at their junction where xylem and phloem tissues were newly developed in unorganized manners, resulting in incidental connection for xylem, but seemingly not for phloem. In these grafts, both apoplasmic and symplasmic transports were achieved. Also the transports of proteins and mRNAs were detected. TEM analysis evidenced the *de novo* secondary plasmodesmata formation at the graft boundaries. Thus, we consider that *Nicotiana* interfamilier grafting, which belongs to incompatible combinations, is functional and practical.

W460: Graft Genetics and Genomics

TBA Evolutionary Impacts of Plant Grafting

Eliezer E. Goldschmidt, The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, P.O.Box 12, Rehovot, Israel

Assessment of the evolutionary significance of plant grafting involves two levels of considerations: Benefits for plant survival (a), and potential role in the formation of new species (b).

(a) The plausible ecological benefits of root grafting have been discussed. Grafting also plays a critical role in the survival of thousands of man-selected plant genotypes which are only propagated by grafting and have not evolved in natural habitats.

(b) Addressing the significance of grafting as a potential evolutionary mechanism must take into account the ancient belief that grafting may give rise to new plant species. Some of the novel species records may be attributed to graft chimeras, which are not true hybrids, but are inheritable. However, recent research has brought us much closer to real evolutionary impacts of grafting, mainly via research by Bock and his team (Germany). Exchange of plastid genome material between stock and scion cells has been demonstrated in interspecific graft junction zones; such changes can become heritable only via lateral shoot formation from the graft site.Subsequently it was shown that entire nuclear genomes are transferred between plant cells in the graft junction zone, leading to the formation of novel, alloploid plant cells. Plants recovered from these cells are real graft hybrids; a fertile, stable hybrid between herbaceous and woody *Nicotiana* species was thus obtained. Similar results have now been obtained with other plant species (Unpubl. data). These findings raise the exciting option that natural grafting may have played an active evolutionary role in plant speciation. It is difficult to estimate the actual participation of this asexual path in plant evolution. Although seemingly dependent upon a rare sequence of events, Nature may have exploited this evolutionary niche as a survival bypass under extreme, catastroph selective pressure conditions.

W461: Graft Genetics and Genomics

Rootstock Scion Somatogenetic Interactions in Perennial Composite Plants

Amit Dhingra, Tyson Koepke and James Crabb, Department of Horticulture, Washington State University, Pullman, WA

In a composite grafted plant, rootstocks control several aspects of scion growth and physiology including yield and quality attributes as well as tolerance to biotic and abiotic stresses. We are interested in understanding the control of floral bud development and overall yield in a perennial crop. Two transcriptome profiling approaches including differential display and 3' untranslated region sequencing was used to identify the genes that are modulated by the rootstock in the developing floral buds of two scion varieties each grafted onto two rootstock genotypes. Differential display identified 207 putatively differentially expressed gene fragments while 3'UTR sequencing and subsequent analyses revealed 115 differentially expressed transcripts. Additionally, near full length transcriptome sequencing was completed to enable clustering and more effective annotations. Blast analysis revealed that several of the differentially expressed genes were transcription factors or DNA binding domain containing proteins which may be involved in altering the number of floral buds.

W462: Graft Genetics and Genomics

Messenger RNA Exchange between Scions and Rootstocks in Grafted Grapevines

Yingzhen Yang[1], Linyong Mao[2,3], Yingyos Jittayasothorn[1,4], Youngmin Kang[1,5], Chen Jiao[2], Zhangjun Fei[2] and Gan-Yuan Zhong[6], (1)USDA-ARS Grape Genetics Research Unit, Geneva, NY, (2)Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY, (3)Department of Biochemistry & Molecular Biology, Howard University, Washington, DC, (4)National Eye Institute, National Institutes of Health, Bethesda,, MD, (5)K-herb Research Center, Korea Institute of Oriental Medicine, Deajeon, South Korea, (6)USDA-ARS, Geneva, NY

Grafting has been widely practiced for centuries in the propagation and production of many vegetable and fruit species. However, the underlying molecular and genetic mechanisms for how the graft partners interact with each other to produce a successful graft remain largely unknown. Using diagnostic SNPs derived from high throughput genome sequencing, we characterized the patterns of genome-wide mRNA exchanges across graft junctions in grafted grapevines grown in the *in vitro* and field conditions. More than three thousand genes transporting mRNAs across graft junctions were identified in this study. These genes were involved in diverse biological processes and many of those which were involved in basic cellular, biosynthetic, catabolic, and metabolic activities, as well as responses to stress and signal transduction, were highly enriched. Field-grown mature grafts had much fewer genes transmitting mRNAs than the *in vitro* young grafts. These mobile mRNAs could move directionally or bi-directionally between scions and rootstocks. The mRNA transmission rates of these genes were generally low, with 65% or more having transmission rates lower than 0.01. Furthermore, genotypes, graft combinations and growth environments had impact on the directions of mRNA movement as well as the numbers and species of mRNAs being exchanged. Moreover, we found evidence for the presences of both passive and selective mechanisms underlying long distance mRNA trafficking in grafted grapevines.

W463: Graft Genetics and Genomics

Xylomic Traits for Rootstock x Scion x Environment Interactions in Tomato

Francisco Pérez-Alfocea, Cristina Martínez-Andújar and Alfonso Albacete, Department of Plant Nutrition, CEBAS-CSIC, E-30100 Murcia, Spain

Grafting onto rootstocks (R) is a surgical technique that enables greater adaptability of elite scion (S) varieties to adverse environments (E). Although this is an ancient agricultural technique that is commercially exploited in several crops, the implementation of suitable RxSxE combinations in Solanaceous, Cucurbitaceous and other high-value vegetable crops represents an untapped opportunity to secure yield stability and reliability under biotic and abiotic stresses (Albacete et al., 2015). However, the current practice only scratches the surface of the vast genetic diversity available, while there is a need, an opportunity and a market for rootstock breeding and use. New useful genetic variations can be introduced as R, but also new genotype interactions can be generated in RxS combinations in a range of stressful environments. Information from gene expression, metabolomics in the xylem (*xylomics*, with hormones playing an essential role) and epigenetic studies, identifying the underlying mechanistic bases of the new phenotypes, will provide new tools to improve identification and prediction of new variants to improve yield, yield stability and quality. Indeed, by using high throughput U-HPLC-MS technology, about 800 metabolites in a range of *m/z* 90-500 have been detected in the root xylem sap of grafted tomato plants, identifying rootstock-genotype specific metabolomic fingerprinting. The number of those metabolites with a biological effect in the scion needs to be addressed. Genetic markers of those traits can be identified from recombinant inbred populations through rootstock phenotyping, QTL and fine mapping analyses.

W464: Graft Genetics and Genomics

Morphological, Physiological and Gene Expression Modulation of Apple Scions By Apple Rootstocks

Gennaro Fazio, USDA-ARS, Geneva, NY and Thomas Tworkoski, USDA ARS Appalachian Fruit Research Station, Kearneysville, WV

Apple scion varieties have been grafted on dwarfing apple rootstocks for several hundred years to increase productivity and early bearing of trees. Recently, we have been able to describe several other rootstock-induced morphological and physiological effects beyond dwarfing and early bearing including branch angle modification, induction of sylleptic branching, nutrient aquisition, reduction in chilling requirements. Apple rootsocks were the first crop where rootstock-induced gene expression modulation in scions was described. We describe a series of experiments aimed at understanding the genetic basis of these rootstock induced changes leveraging breeding populations, QTL and eQTL approaches and phenotyping of composite trees.

W465: Gramene Project

Exploring and Comparing Plant Genomes Using the Gramene/Ensembl Plants Browser

Joshua Stein[1], Sharon Wei[1], Marcela Karey Tello-Ruiz[1], Andrew Olson[1], Sunita Kumari[2], Joseph Mulvaney[1], Jim Thomason[2], Dan Bolser[3], Arnaud Kerhornou[3], Paul J. Kersey[3] and Doreen Ware[4], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)Cold Spring Harbor Laboratory, NY, NY, (3)EMBL - The European Bioinformatics Institute, Cambridge, United Kingdom, (4)Cold Spring Harbor Laboratory/USDA-ARS, NY, NY

Learn the benefits of using Gramene and Ensembl Plants (www.gramene.org & www.plants.ensembl.org) to accelerate your research goals. We host genome browsers for 39 complete reference genomes, each displaying value-added annotations, gene-trees, and whole genome alignments. Evolutionary histories are provided in phylogenetic gene trees that classify orthologous and paralogous relationships as speciation and duplication events. Orthologous genes inform synteny maps that enable interspecies browsing across ancestral regions. Browsers from multiple species can be viewed simultaneously, with links showing homologous gene and whole-genome alignment mappings. SNP and structural diversity data, available for eleven species, are displayed in the context of gene annotation, along with the consequence of variation on transcript structure (e.g. Missense variant). For each variant, drill-down to individual accession genotypes within the study's diversity panel. New data include the maize methylome and RNA-seq for each of ten Oryza species. New BLAST service displays alignment results directly onto genome browsers and karyotype views. Privately upload your own experimental data, such as RNA-seq alignments, to view in the context of gene annotations. All reference assemblies are synced with iPlant Genome Services giving cross-platform compatibility between your analyses and Gramene displays. Download visual displays as high-resolution, publication-ready, image files, along with the underlying supporting data in conventional file formats. Mine data using Gramene's BioMart; perform complex queries of annotation, homology and variation data, then click to visualize results. Gramene is funded by a grant from the NSF (IOS-1127112).

W466: Gramene Project

Plant Reactome: A Resource for Comparative Plant Pathway Analysis

Pankaj Jaiswal, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR

The Plant Reactome database (http://plantreactome.gramene.org/) developed collaboratively by Gramene and the human Reactome project hosts metabolic, genetic and signaling pathways for several model and crop plant species. The Reactome data model organizes gene products, small molecules and macromolecular interactions into reactions and pathways in context of their subcellular location to build a systems-level framework of an eukaryotic cell. The Plant Reactome features *Oryza sativa* (rice) as a reference species that was built by importing RiceCyc metabolic network; adding highly conserved projected pathways (e.g., cell cycle, DNA replication, transcription, translation, etc.) from human Reactome; and curating new metabolic, signaling and genetic pathways. The Plant Reactome database now contains 238 rice reference pathways and orthology-based pathway projections for 33 plant species. Plant Reactome allows users to i) compare pathways across various plant species; ii) query and visualize curated baseline and differential expression data available in the EMBL-EBI's Expression Atlas in context of pathways listed in the Plant Reactome; and iii) analyze genome-scale expression data and conduct pathway enrichment analysis to identify pathways likely to be modulated in response to environmental stresses and experimental treatments. Plant Reactome data links to gene loci pages in the genome section of the Gramene and to external sources like, UniProt, PIR, ChEBI, PubChem, PubMed, and GO. Users can access/download our data in various formats from our web site and via APIs. The presentation will discuss tools for pathway enrichment analysis and homologue pathway comparison, development of the Plant Reactome portal, curation of reference rice pathways, and phylogeny based analyses of projected pathway annotations. The Gramene database project is supported by an NSF award (IOS-1127112). Intellectual and infrastructure support for the Plant Reactome is provided by the Human Reactome award (NIH: P41 HG003751, ENFIN LSHG-CT-2005-518254, Ontario Research Fund, and EBI Industry Programme).

W467: Gramene Project

Exploring Gramene's Comparative Genomics Datasets Using New Search Tools

Joseph Mulvaney[1], Andrew Olson[1], Jim Thomason[2] and Doreen Ware[3], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)Cold Spring Harbor Laboratory, NY, NY, (3)Cold Spring Harbor Laboratory/USDA-ARS, NY, NY

Gramene is a unique comparative genomics resource that comprises genomic, pathway, diversity, and expression databases for economically important species across the plant kingdom.

I am excited to introduce and demonstrate a new search interface (http://search.gramene.org) that integrates data from these databases and helps users search, interpret and explore our data. You will see how the application proactively suggests appropriate terms as you type; shows an interactive distribution of result sets across genomes and species; provides statistics and enriched terms in the result set; and allows users to use details of a result to expand or narrow down a search.

The search interface uses a public web service (http://data.gramene.org) that is well documented and available for access from custom scripts. I will demonstrate how to construct simple queries and receive the results either in CSV or JSON format.

These tools are in active development and we would love to hear what you think about them and how you think they can be improved.

W468: Gramene Project

The European Variation Archive: A Central Repository for Plant Genetic Variation Data

Gary Saunders, EMBL-EBI, Cambridge, United Kingdom

In collaboration with Gramene and Ensembl Plants, the European Variation Archive (EVA; www.ebi.ac.uk/eva) aims to provide a resource to the plant community that can accession, archive and provide views of genetic variation data more quickly and in a more granular manner than any other resource, worldwide. EVA works with data submitters to archive genetic variation data, in Variant Call Format (VCF) files, to confirm that archived data is truly valid and associated with rich metadata. This ensures that these data are of most benefit to the community and we shall present an overview of this submission process.

At EVA, archived variants are normalized and annotated using a variety of standardized methods, including Ensembl's Variant Effect Predictor. We shall present the ways in which users can mine these data using filters on the EVA website to construct both study-centric and global queries, filtering on any combination of species, methodology, variant type, phenotype, consequence or allele frequency and show how results from these queries can be downloaded in a variety of formats including VCF and CSV. Additionally, EVA provides a comprehensive RESTful web-service, to allow programmatic access, and hence the integration of these data with other resources (such as Ensembl Plants and Gramene) and this shall also be presented.

W469: Gramene Project

Developments for Plant Data in Expression Atlas

Robert Petryszak, EMBL - The European Bioinformatics Institute, Hinxton, United Kingdom

Expression Atlas (http://www.ebi.ac.uk/gxa) provides information about gene and protein expression in animal and plant samples of *e.g.* cell types, organism parts, developmental stages, strains and disease. The number of plant studies and species in Atlas has doubled in the last year - to over 430 studies (and 14 plant species) to date. Atlas is now backed by extensive RNA-seq analysis of all public plant data - resulting in gene expression data from over 10,000 high-quality 'runs' (600 studies) in 44 species, that are now being steadily curated and loaded into Atlas. This talk will provide overview of the new data, highlight our current work on novel analyses of Atlas data (*e.g.* gene co-expression) and present our plans for the coming year.

W470: Gramene Project

Improving Reference Genome Resources Using Long-Read Sequencing Technology

Bo Wang[1], Elizabeth Tseng[2], Michael Regulski[1], Tyson A. Clark[2], Ting Hon[2], Yinping Jiao[3], Jerry Lu[1], Andrew Olson[1], Joshua Stein[1] and Doreen Ware[4], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)Pacific Biosciences, Menlo Park, CA, (3)USDA-ARS/Cold Spring Harbor Laboratory, Lubbock, TX, (4)Cold Spring Harbor Laboratory/USDA-ARS, NY, NY

Zea mays is an important crop species and genetic model for elucidating transcriptional networks in plants. Uncertainties about the complete structure of mRNA transcripts, particularly with respect to alternatively spliced isoforms, limit the progress of research in this system. In this study, we used single-molecule sequencing technology to investigate the maize transcriptome. Intact full-length cDNAs from six tissues of the maize inbred line B73 were barcoded, pooled, size-fractionated (<1 kb, 1–2 kb, 2–3 kb, 3–5 kb, 4–6 kb, and 5–10 kb), and sequenced on the PacBio RS II platform with P6-C4 chemistry. The resultant 111,151 transcripts captured ~70% of the annotated genes of the maize RefGenV3 genome assembly. A large proportion of transcripts (57%) represented novel, sometimes tissue-specific, isoforms of known genes, and 3% corresponded to novel gene loci. In other cases, the identified transcripts have improved existing gene models. To validate transcript structures we checked for occurrence of each splice-junction within high-depth Illumina reads generated from matched tissues. Averaging across all six tissues, 90% of splice-junctions were well supported by short-reads in matched tissues. In addition, we identified a large number of novel long non-coding RNAs (lncRNAs) and fusion transcripts, and found that DNA methylation plays important roles in generation of various isoforms. Our results show that the characterization of the maize B73 transcriptome is far from complete, and that maize gene expression is more complex than previously thought.

W471: Grape Genome Initiative

Field Systems Biology of Grapevine Diseases

Dario Cantu, Abraham Morales-Cruz, Katherine C.H. Amrine, Rosa Figueroa-Balderas, Gabrielle Allenbeck and Barbara Blanco-Ulate, University of California Davis, Davis, CA

Vitis vinifera cultivars are susceptible to a wide range of pathogens that infect vegetative, woody and fruit tissues. Understanding how grapevines and pathogens interact in the vineyard is crucial to improve crop traits under naturally fluctuating environments. In the last three years we have developed genome and transcriptome information for some of the most important grapevine fungal pathogens, including the causal agents of

powdery mildew and trunk cankers. Besides the annotation of potential virulence factors, these genomes have provided the references necessary to determine fungal species composition and virulence factor expression patterns in diseased vineyard samples using metagenomics and metatranscriptomics approaches. We have also applied systems biology approaches to characterize changes in grapevine development and responses to biotic stress under field conditions. We have integrated transcriptomics, metabolomics, and enzyme activity assays to characterize the impact of noble rot on the development and metabolism of white-skinned grape berries. We identified grape molecular pathways that were triggered by *Botrytis cinerea* during noble rot, some of which correspond to plant stress responses, while others are related to ripening processes. Developmental and metabolic changes induced by noble rot also occur during normal development of red-skinned berries, including the activation of secondary metabolic pathways that are absent or have limited flux in white-skinned grape cultivars. We further validated the induction of these pathways using noble-rotted berries from two additional years. We also confirmed that the impact of noble rot on grape metabolism is still detectable in botrytized wines produced from the same vineyard.

W472: Grape Genome Initiative

Transcriptional Regulation of Grapevine Bud Burst in Light and Dark Conditions

Sandra Patricia Agudelo-Romero[1], Karlia Meitha[1], Oliver Berkowitz[1], Colin S Gordon[2], Christine H Foyer[3], John A Considine[1] and Michael J Considine[4], (1)University of Western Australia, Crawley, Australia, (2)Department of Agriculture and Food Western Australia, South Perth, Australia, (3)University of Leeds, Leeds, United Kingdom, (4)University of Western Australia, Perth, Australia

The transition from quiescence to bud burst is rapid, taking 3-5 days from completely enclosed bud to a fully open and photosynthetic organ. However, little is known about what process underlining during grapevine buds transition. Here, we studied a developmental time course of bud burst grown in a light/dark photoperiod (0h, 24h, 72h and 144h) and continuous darkness (0h, 3h 24h, 72h and 144h) to gain knowledge about the processes involved in grapevine bud burst (Crimson Seedless).

To investigate the requirement for light to enable viable bud burst, RNA-Seq experiment was done (FDR 0.01 and FC 3-fold). In both, light/dark photoperiod and continuous darkness were performed comparisons against 0h. Functional enrichment (FatiGO, *P*=0.01) was performed to each treatment.

Categories as cellular process, metabolism (cellular, primary and secondary), signaling, response to stimulus and transport overview were altered between treatments. Genes belonging to NAC family transcription were detected down-regulated whole time course in continuous darkness, whereas ERF genes related to AP2 family transcription factor were found in light/dark photoperiod. Remarkably, Circadian clock Signaling category was down-regulated in the early stages with a slight increased at 144h in both treatments. However, different genes built their profiles. Furthermore, C2C2-YABBY family transcription factor was detected up-regulated in light/dark treatment.

Altogether the data provides an overview of gene expression alteration due to light exposition in grapevine buds and how these treatments affected the initiation of bud burst. Suggesting a pivotal role of several functional categories as metabolism (primary and secondary) as well as transport overview.

W473: Grape Genome Initiative

A Phocus on Phenotyping: Opportunities and Challenges in Local and Centralized Trait Evaluation from the *Vitis*Gen Experience

Lance Cadle-Davidson[1], Gavin Sacks[2], Anne Fennell[3], David Gadoury[4], Qi Sun[5], Peter Schweitzer[2], Jason Londo[1], Craig A. Ledbetter[6], Matthew D. Clark[7], James J. Luby[7], Peter Hemstad[7], Adrian Hegeman[7], Soon Li Teh[7], David Manns[4], Paola Barba[4], Chin-Feng Hwang[8], Surya Sapkota[8], Li-Ling Chen[8], Jonathan Fresnedo Ramirez[9], Shanshan Yang[10], Elizabeth M. Takacs[10] and Bruce Reisch[10], (1)USDA-ARS Grape Genetics Research Unit, Geneva, NY, (2)Cornell University, Ithaca, NY, (3)South Dakota State University, Brookings, SD, (4)Cornell University, Geneva, NY, (5)Bioinformatics Facility, Cornell University, Ithaca, NY, (6)Crop Diseases, Pests and Genetics Research Unit, USDA-ARS, Parlier, CA, (7)University of Minnesota, St. Paul, MN, (8)Missouri State University, Mountain Grove, MO, (9)Institute of Biotechnology, Cornell University, Ithaca, NY, (10)School of Integrative Plant Science, Cornell University, Geneva, NY

The integration of relevant genetic resources, robust phenotypes, and cutting-edge genotypic data is a challenge that individual scientists rarely overcome successfully. In the USDA-NIFA *Vitis*Gen project (www.vitisgen.org) for grapevine cultivar improvement, our research team has pursued a shared strategy 1) to maintain breeding germplasm locally, 2) to collect and analyze high-resolution genotypic data centrally, and 3) to collect and analyze phenotypic data both locally and in phenotyping centers specializing in fruit chemistry, low temperature response, and powdery mildew resistance. In the process of genotyping 28,000 individuals and collecting millions of phenotypic data points, challenges and unexpected opportunities arose from our approach. Here, we will 'Phocus' on Phenotyping, sharing our experiences and highlighting what seems to be working and not working in both local and central trait evaluation, based primarily on genetic analysis. These stories and results may facilitate both cultivar improvement and strategies for genetic analysis of traits.

W474: Grape Genome Initiative

***De novo* Genome Assembly of Heterozygous *Vitis* species Using Next-Generation Sequencing**

Sagar Patel[1], Padmapriya Swaminathan[1], Erliang Zeng[2] and Anne Fennell[1], (1)South Dakota State University, Brookings, SD, (2)University of South Dakota, Vermillion, SD

Vitis vinifera cultivars are widely used for wine, table and raisin production throughout the world. A reference genome for the *V. vinifera* inbred line (PN40024, ~93% homozygous) is available; however, standard cultivars are highly heterozygous. We evaluated assemblers ALLPATHS-LG and PLATANUS for the assembly of heterozygous genomes using next-generation derived sequences. *V. vinifera* 'Sultanina' reads were downloaded from NCBI (accession no SRP026420) and *V. riparia* (PI588259) reads were generated from our laboratory. *De novo* assemblies were developed for the two heterozygous genotypes and the structure and quality of the assemblies were assessed in relation to *V. vinifera* (PN40024, V2). The Next-Gen sequence library diversity and assembler type influenced the quality of assembly for highly heterozygous species.

W475: Grape Genome Initiative

The Transcriptome of Quiescence and Dormancy in Subtropical and Mediterranean Grapevine

Michael J Considine[1], Sandra Patricia Agudelo-Romero[1], John A Considine[1], Oliver Berkowitz[1,2] and Colin S Gordon[3], (1)University of Western Australia, Crawley, Australia, (2)La Trobe University, Bundoora, Australia, (3)Department of Agriculture and Food Western Australia, South Perth, Australia

Vine physiology is disorderly and requires intensive chemical and physical management, as the seasonal cues that grapevine relies on for developmental transitions are lacking. To gain insight into how differences in the temperature due to climate features can modify grapevine bud dormancy, a RNA-seq study was performed to investigate differences between subtropical and Mediterranean climates in table grapes (Flame Seedless).

For this, gene expression changes in buds from two adjacent vineyards in subtropical Western Australia (25°S latitude) were compared against one vineyard in a Mediterranean climate (32°S). Buds were collected at the end of summer and in mid-winter, over two successive years (2012 and 2013). Principal Components Analysis (PCA) revealed that the main factor explaining the global gene expression differences was the seasonality.

Differential expression analysis of subtropical and Mediterranean climates were performed (i) at the beginning and (ii) at the end of grape bud dormancy, and functional enrichments were carried. The first comparison showed a large difference between two consecutive seasons, while the second comparison showed major changes between climates. WRKY family transcription and oxidative stress (Glutathione S-transferase) categories sowed differences between climates at the beginning of dormancy. Whereas at the end was detected ethylene-mediated Signaling pathway and C2C2-YABBY family transcription factor categories.

This work provides a global view of major transcriptional changes taking place in Australian subtropical and Mediterranean climates, highlighting those molecular and biological functions that showed differences between climates, suggesting a main role of those functional categories during regulation of bud dormancy.

W476: Grape Genome Initiative

An Update on Sequencing the Cabernet Sauvignon Genome / Towards a Grapevine Information System

Grant R. Cramer[1], Ryan Ghan[1], Alberto Ferrarini[2], Rosa Figueroa-Balderas[3], Andrea Minio[2], Massimo Delledonne[2], Dario Cantu[3] and Anne-Francoise Adam-Blondon[4], (1)University of Nevada, Reno, Reno, NV, (2)University of Verona, Verona, Italy, (3)University of California Davis, Davis, CA, (4)INRA - URGI, Versailles, France

Sequencing of Clone 8 of Cabernet Sauvignon (*Vitis vinifera L.*) was initiated in 2011. Several different stages with different technologies have taken place since then. At an earlier stage, an assembly of the genome was attempted with approximately 110x Illumina reads and 5x PacBio reads. Different assemblies were compared for denovo assembly using SOAPdenovo2, and with a hybrid approach using SPAdes V3.0, Celera WGS8.1, and SSPace-longreads. The longer sequencing reads from PacBio made major improvements in the assembly compared to the SOAPdenovo2 results with Illumina reads only. However, the assembly results were still unsatisfactory, so an additional 100x genome coverage with PacBio RS II chemistry have been generated. An update on the current sequencing results and status of the assembly will be presented./In the frame of the IGGP, a strategy for the development of a grapevine international information system has been achieved and has now to be implemented into a road map. A presentation of the strategy will be made for discussion with a larger community.

W477: Grass Genome Initiative (IGGI)

Introduction

Katrien M. Devos, Institute of Plant Breeding, Genetics and Genomics (Dept. of Crop and Soil Sciences), and Dept. of Plant Biology, University of Georgia, Athens, GA

W478: Grass Genome Initiative (IGGI)

The Ancient Genome of an Aquatic Plant, *Spirodela polyrhiza,* at the Root of Monocot Evolution

Wenqin Wang, Waksman Institute, Rutgers University, Piscataway, NJ, Paul Fourounjian, Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ and **Joachim Messing**, Rutgers University, Piscataway, NJ

Whole-genome sequencing has been critical in advancing our knowledge of the dynamics of ancient chromosomal rearrangements during speciation. Because of their agronomic significance the scientific community has given most attention to the family of species that contribute most to calorie production in the world, the grasses or gramineae. The grasses belong to the monocotyledonous division of flowering plants. To better understand the evolution of monocots, it has become necessary to sequence the genome of species outside the order of the Poales that comprises the gramineae. The Greater Duckweed or *Spirodela polyrhiza*, which had been sequenced by an international consortium (Nat. Commun. e3311), offers a unique perspective to study this evolution. As a member of the Alismatales, *S. polyrhiza* represents a well-preserved genome of a distant basal monocot. Developmentally and morphologically, Spirodela is very distinct from the gramineae, due to its aquatic, clonal lifestyle, and rare irregular flowering. This distinct genome therefore represents a strategic perspective to view traits that have been conserved through this phylogenetic distance, as well as those that have been altered for this distinct lifestyle.

W479: Grass Genome Initiative (IGGI)

DNA Methylation in the Grasses: Through the Lens of Comparative Epigenomics

Chad E. Niederhuth, University of Georgia, Athens, GA

DNA methylation is a major epigenomic modification that is involved in a host of processes from transposon silencing, gene expression, and chromatin structure. In plants, it can also be inherited across generation and differences in methylation can lead to heritable phenotypic consequences without changes to the underlying DNA sequence. We used MethylC-seq and a comparative epigenomics approach to study DNA methylation across 34 Angiosperm species, many of which do not have previously published methylomes. We found extensive variation in

methylation between species in all contexts. By taking into account phylogenetic relationships, we are able to identify lineage specific differences in methylation and how these relate to the evolution and origin of genes in each species.

W480: Grass Genome Initiative (IGGI)

The Alpha and Omega of Sigma and Tau; Paleopolyploidy in the Monocots

David Sankoff and Chunfang Zheng, University of Ottawa, Ottawa, ON, Canada

The structure of the genome ancestral to the Poaceae (Poales) has been controversial, obscured by three rounds (τ, σ and ϱ) of whole genome duplication (WGD) since the emergence of the monocots, each WGD followed by a period of fractionation and rearrangement. Pineapple, also in the order Poales, but in the basal bromeliad family, escaped the ϱ event. The recent sequencing and annotation of the pineapple genome allowed us a clearer picture of τ and σ so that we could use computational tools to fully resolve the structure of the pre-σ, pre-τ (commelinid) precursor before it radiated into the Arecales, Poales, Zingiberales and other orders. Using as an outgroup the recently sequenced duckweed genome, a monocot that diverged before τ, but that has undergone two successive WGD in its own lineage, we applied the technique of "augmented synteny blocks" to the pineapple genome to clearly establish a seven-chromosome commelinid ancestor. This validates the conclusion of a seven-chromosome ancestor of the cereal family, rather than the five-chromosome or nine-chromosome constructs that also seemed credible based only on sequences within the Poaceae.

W481: Grass Genome Initiative (IGGI)

Comparative proteomic analysis of ancient plants and monocots identifies proteins important for rhizome growth and development

David R. Gang[1], Fernanda Salvato[2], Ruifeng He[1], William Nelson[3], Min-Jeong Kim[1], Tiago Balbuena[4], R. Shyama Prasad Rao[5], Carol Soderlund[3] and Jay Thelen[6], (1)Institute of Biological Chemistry, Washington State University, Pullman, WA, (2)State University of Campinas, Sao Paulo, Brazil, (3)University of Arizona/Bio5 Institute, Tucson, AZ, (4)Sao Paulo State University “Julio de Mesquita Filho”, Jaboticabal, Brazil, (5)Biostatistics and Bioinformatics Division, Mangalore, India, (6)University of Missouri-Columbia, Columbia, MO

The rhizome was the original plant stem that confers competitiveness and invasiveness to many of the world’s worst weeds, such as the grasses Johnsongrass (*Sorghum halepense*), Bermuda grass (*Cynodon dactylon*), quack grass (*Agropyron repens*), cogon grass (*Imperata cylindrica*), common reed (*Phragmites australis*) and red rice (*Oryza longistaminata*), and primitive plants such as horsetail (*Equisetum arvense*). In addition rhizomes impart positive economic benefits in plants such as ginger (*Zingiber officinale*), turmeric (*Curcuma longa*), miscanthus (*Miscanthus x giganteus*) and giant reed (*Arundo donax*). An extensive comparative transcriptomic, proteomic and metabolomic investigation of these plants has identified a number of genes, metabolites and proteins that are rhizome-specific and that appear to play important roles in regulating rhizome differentiation and growth. Although integration of all datasets will be presented, the focus of this talk will be on the proteins identified, and their putative or confirmed roles in rhizome growth and development.

W482: Grass Genome Initiative (IGGI)

A Solution to the C-Value Paradox and the Function of Junk DNA

Michael R Freeling, University of California, Berkeley, CA

Plants exhibit the **C-value paradox** more dramatically than do animals: related genomes with approximately the same number of chromosomes, number of genes and a similar degree of “complexity” differ significantly in genomic DNA content. The DNA difference is often called “junk” because – whether it functions or not—this DNA is proved to be “excess”. An experimentally useful edition of the C-value paradox exists within the maize genome itself, as will be explained. The official origin of this talk has been published[1]. We think junk DNA --composed of dead transposons-- functions in bulk as ballast, both in quantitative gene regulation and in coordinating movement of chromosomes at anaphase. Tests will be discussed. 1. Freeling M, Xu J, Woodhouse M, Lisch D: **A Solution to the C-Value Paradox and the Function of Junk DNA: The Genome Balance Hypothesis**. *Mol Plant* 2015. 8: 899-910.

W483: Grasslands (Lolium Genome Initiative)

Genomic Dissection and Prediction of Heading Date in Perennial Ryegrass

Dario Fè[1], Fabio Cericola[2], Stephen Byrne[3], Ingo Lenk[1], Bilal H. Ashraf[2], Morten Greve-Pedersen[1], Niels Roulund[1], Torben Asp[3], Luc L. Janss[2], Christian Sig Jensen[1] and Just Jensen[2], (1)DLF A/S, Store Heddinge, Denmark, (2)Molecular Biology and Genetics, Aarhus University, Tjele, Denmark, (3)Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark

So far, the implementation of Genomic Prediction (GP) in plant breeding has been mainly investigated in crops farmed as genetically homogeneous cultivars. The aim of this work is to test the feasibility of GP in forage perennial ryegrass (*Lolium perenne* L.), using breeding material from different steps of the breeding program.

Heading date was used as a model trait, due to its ease of assessment and relatively high heritability. Genomic data was produced by Genotyping-By-Sequencing of family bulk samples, and mean allele frequencies were estimated per family at each marker. Phenotype data was also collected on family means from sward plots.

Analyses involved: (i) Genome Wide Association study, within F2 families, in order to uncover the genetic architecture of the trait; (ii) estimation of the breeding values in a set of related synthetics, performed both by Marker Assisted Selection (MAS) and through GP; (iii) comparison of different models for GP within F2 families, using different cross-validation schemes, designed to test specific hypothesis.

A total 19 markers were significantly associated with heading date, some of them locating within or proximal to genes involved in floral regulation (e.g. CONSTANS and Phytochrome C). Other markers were situated in the intergenic region, revealing a relatively complex genetic architecture. GP models yielded very accurate predictions, significantly superior to those obtained with MAS using the 19 significant markers. Higher accuracies were found when predicting from related families than when predicting families unrelated to families used to train the model.

W484: Grasslands (Lolium Genome Initiative)

A Composite Approach to Study Drought Tolerance in Tall Fescue

Malay C. Saha[1], Shyamal K Talukder[1], Perumal Azhaguvel[2], Shreyartha Mukherjee[1] and Yuhong Tang[1], (1)The Samuel Roberts Noble Foundation, Ardmore, OK, (2)Syngenta, Slater, IA

Tall fescue is a cool-season perennial grass grown in the temperate regions of the world. The forage-purpose tall fescue belongs under two distinct morphotypes, the Continental and the Mediterranean. Drought is the single most constrain for tall fescue cultivation in the southern Great Plains. Enhancing drought tolerance is the key strategy for improving persistence of Continental fescue in the region. A suitable drought screening protocol was developed in our program. A transcriptome profiling between water-stress tolerant and susceptible genotypes unraveled the genetic regulatory mechanism of water-stress responses. A total of 199,399 contigs were assembled with an average length of 585 bp. Significantly differentially expressed reference transcripts (RTs, 2986) were identified and 1048 of them could be annotated into functional groups. In total 175 RTs were reported for various stress-related functions and 65 of them were encoded kinase proteins and 40 each were encoded transposons and transporter proteins. Summer dormancy is an important drought adaptation trait, which can improve persistence of the Mediterranean fescue. Screening experiments suggested that physiological traits, i.e., stomatal conductance, can be used to discriminate summer dormancy. Multi-factor transcriptome sequencing has been conducted using leaf, root and crown tissues from genotypes of both morphotypes to understand and identify the genetic mechanism of summer dormancy. A de novo assembly assembled 710,125,152 bp sequences into contigs. A total of 1,031,513 RTs were identified in 583,736 contigs. Differential transcriptional responses were documented among the genotypes, tissue types and three time points. Further exploration to pin down the background mechanism is undergoing.

W485: Grasslands (Lolium Genome Initiative)

Perennial Ryegrass Genome Draft Assembly and Anchoring by GBS

Ewan Mollison[1], Janaki Velmurugan[2], Dan Milbourne[3], Susanne Barth[4], David Marshall[1] and Linda Milne[1], (1)The James Hutton Institute, Dundee, United Kingdom, (2)Teagasc CELUP, Carlow, Ireland, (3)Crops, Environment & Land Use Programme, Teagasc, Carlow, Ireland, (4)Teagasc, Carlow, Ireland

We describe a genotyping by sequencing (GBS) study in perennial ryegrass, *Lolium perenne*, and its usage in anchoring scaffolds from a *de novo* assembled genome of the same.

A draft *de novo* genome assembly for *L. perenne* was generated using Illumina sequencing of paired-end (PE), mate-pair (MP) and long jumping distance (LJD) libraries. Using this, some 207-fold raw coverage was achieved, reduced to around 105-fold by subsequent quality control. The draft assembly contains 424,750 scaffolds, reflecting 1.11Gbp of the 2.5Gbp genome, with scaffold N50 of 25,213 bp and contig N50 of 3,790 bp.

Gene regions were identified using a combination of *ab initio* gene prediction, in-house RNA-Seq, comparison with publicly available *Lolium* transcriptome sequences, and homology with related species. Comparing gene-containing scaffolds identified by these methods resulted in a core set of 12,245 scaffolds that were found to be common to all four sets, equivalent to around 482 Mbp in length.

In conjunction with *de novo* genome assembly, a high-density SNP-based genetic linkage map, with 3,092 SNPs and 7,260 presence/absence variants (PAVs), was developed from an F2 Biomass mapping population of *L. perenne*. This map was used to anchor 4,767 genomic scaffolds with a total size of approximately 200Mbp, or 18% of the assembled genome size.

W486: Grasslands (Lolium Genome Initiative)

Breeding and Genomic Resources for Intermediate Wheatgrass and Perennial Agriculture

Steve Larson[1], James A Anderson[2], Xiaofei Zhang[2], Kevin M. Dorn[2], Jesse Poland[3], Lee DeHaan[4], Traci R. Kantarski[5], Kevin Jensen[1] and Blair Waldron[1], (1)USDA-ARS Forage and Range Research, Logan, UT, (2)University of Minnesota, St. Paul, MN, (3)Kansas State University, Manhattan, KS, (4)The Land Institute, Salina, KS, (5)Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, Manhattan, KS

Intermediate wheatgrass (*Thinopyrum intermedium*) is a cool-season perennial grass used for soil conservation, grazing, and hay production, and has demonstrated potential for the development of perennial grain and biomass crops. As a tertiary source of disease resistance genes, intermediate wheatgrass (IWG) ranks among closest perennial relatives of wheat (*Triticum aestivum*). Like wheat, IWG is allohexploid ($2n=6x=42$) and has a geographic origin, center of diversity, and adaptation range similar to wheat. A survey of 182 IWG accessions reveals two genetically distinct and diverse meta-populations, spanning Europe and Asia, which provide a valuable source of natural germplasm. Moreover, advanced genomic resources will be used to elucidate genetic differences between annual and perennial relatives of wheat and develop perennial grain, forage, and biomass crops. Molecular-genetic maps have been constructed from full-sib IWG populations, using genotype-by-sequencing (GBS) approaches, with a goal of constructing high-density maps for QTL mapping, genome sequence assembly, and genomic selection. Genome and transcriptome reference sequences from a haploid IWG plant (estimated 1C genome size of 12.6 Gb), and related population, will be assembled using a POPSEQ approach with a goal of 90X coverage from short insert paired-end and large insert mate-pair libraries. Field evaluations have been established to identify genes and QTLs controlling functionally-important seed, grain, and biomass production traits in diverse environments and management regimes including non-competitive and competitive grass-legume species mixtures. Comparisons of gene expression profiles from various tissues of annual and perennial forms of wheat, wheatgrass, and *Brachypodium* aim to elucidate the genetic control of perennial regrowth.

W487: Grasslands (Lolium Genome Initiative)

Application of Historical Data from Commercial Ryegrass Breeding to Enable Rapid Implementation of Genomic Selection

Luke Pembleton[1], Courtney Inch[2], Hiroshi Shinozuka[1], Bec Baillie[1], Michelle Drayton[1], Preeti Verma[1], German C. Spangenberg[3], Hans D. Daetwyler[1] and Noel O.I. Cogan[1], (1)Department of Economic Development, Jobs, Transport & Resources, Bundoora, Australia, (2)New Zealand Agriseeds, Christchurch, New Zealand, (3)Department of Economic Development, Jobs, Transport & Resources, Melbourne, Australia

The application of genome-wide molecular markers and quantitative genetics to predict plant performance will change the way pasture plants are bred. Genomic selection offers the ability to predict the performance of selection candidates with only genotypes and marker effects estimated in

a reference population comprised of genotyped and phenotyped individuals or populations. The main factors affecting the accuracy of genomic selection are reference population size, trait heritability, and breeding population genetic diversity. The global perennial ryegrass population is diverse and would need a large reference population to achieve moderate accuracy of genomic selection. However, the diversity within a breeding program may be less, which allowed us to apply genomic selection across cultivars/varieties with moderate reference size. We combined historical phenotype records for seasonal biomass yield and heading date from 15 years of commercial perennial ryegrass breeding with high-density population based transcriptome genotyping-by-sequencing. The ability to genomically predict the observed phenotypic performance in each progressive year was explored by using all varieties from previous years as a reference population and applying a range a genomic selection equations. Accuracies achieved for biomass and heading date were moderate to high, respectively, reflecting the relative heritabilities of these two agronomic traits. These results highlight the ability to predict sward-based phenotypic performance early in the breeding program of individual plants, rapidly reducing the breeding cycle and consequently at least doubling the rate of genetic gain. This study along with the commercially relevant reference population now enables rapid adoption and implementation of genomic selection in ryegrass.

W488: Grasslands (Lolium Genome Initiative)

Selection and Genome-Wide Prediction of Phenology and Biomass Yield in Miscanthus

Chris Davey[1], Rick Nipper[2], Paul R. H. Robson[3], Kerrie Farrar[4], John C. Clifton-Brown[4], Elaine Jensen[4], Iain Donnison[5] and Gancho T. Slavov[4], (1)IBERS- Aberystwyth University, Aberystwyth, United Kingdom, (2)Floragenex Inc., Eugene, OR, (3)IBERS - Aberystwyth University, Aberystwyth, United Kingdom, (4)Aberystwyth University, Aberystwyth, United Kingdom, (5)Aberystwyth University, Aberystwyth University, United Kingdom

Miscanthus has great potential as a biofuel crop because of its high productivity for low input. However, it is an undomesticated out-breeding perennial which makes the rapid development of commercial varieties demanded for carbon mitigation a challenging process. Therefore, the aim of this study was to access the potential of genomic selection (GS) to predict biomass yield and traits correlated to it, using a genome wide marker set. The GS predictive abilities (correlation of predicted and observed genetic values) across traits ranged from 0.76 for day of flag leaf production down to 0.06 for yield itself. To improve the yield prediction, a set of selection indices were built using between one and eight additional traits, aiming to minimise the phenotyping effort needed by breeding programs, while jointly maximising the response to selection and predictive ability through GS. The relative response to selection on the indices compared to yield alone were up to 16% higher, whereas GS on the selection indices resulted in predictive abilities that were up to six times higher than for yield. This is a substantial increase in the ability to make predictions about biomass yield and with further work could be used, for instance, to prioritise crosses, as well as to select seedlings directly from their marker profiles hence reducing breeding-cycle times.

W489: Host-Microbe Interactions

The Making of Tree Pathogens: A Genomic Dissection

Richard C. Hamelin, University of British Columbia, Vancouver, Canada

Tree pathogens cause some of the most devastating outbreaks in forests. In spite of their importance, we know relatively little about their origin and sources and the basic common characteristics of tree pathogens. We applied a genomics approach to understand how the poplar canker pathogen, caused by

Mycosphaerella populorum (anamorph: *Septoria musiva* syn. *Sphaerulina musiva*), evolved to cause one of the most severe diseases of poplars across various hosts and landscapes. This pathogen has recently expanded geographically and has been increased its host range to additional species such as the black cottonwood, *Populus trichocarpa* and the balsam poplar, *P. balsamifera*. Genome comparison between *M. populorum*, which causes both leaf spots and cankers, and *M. populicola*, a close relative that also attacks poplar but does not cause cankers, indicates that *M. populorum* has acquired a unique set of genes via horizontal gene transfer (HGT) that is important to its ability to attack woody hosts. Population genomic analysis of *M. populorum* samples from across its distribution range highlights an unusually high and uneven level of diversity even in recently introduced populations. The patterns suggest both extensive migration and possibly additional unsampled ghost populations that could have contributed to the recent expansions. We hypothesize that a jump to *P. balsamifera* could be driving speciation by imposing selection at the mating type locus, causing a shift from sexual to asexual reproduction, followed by isolation and drift. These results highlight very dynamic mechanisms leading to host colonization and incipient speciation and specialization in this tree pathogen.

W490: Host-Microbe Interactions

Fish Microbiota: The Silent Majority Speaks at Last!

Nicolas Derome, Institut de Biologie Intégrative et des Systèmes (IBIS), Quebec, QC, Canada

Gut microbes outnumber the cells of their hosts by a 10:1 ratio. In humans, those microbes can provide up to 250,000 unique genes that play a role in improving host metabolic and immune functions, for instance. My talk will address the indispensable functions of the microbiota, and more importantly focus on the adaptative potential that microbiota provides to its host. These evolutionary and ecologically relevant aspects will be discussed with special emphasis on the current knowledge of host-microbiota interactions in fish.

W491: Host-Microbe Interactions

Evaluation of Bovine Rumen Microbial Populations Following Periods of Dietary Restriction and Subsequent Re-alimentation

Sinead Waters, Animal and Grassland Research and Innovation Centre, Carlow, Ireland

Authors: Matthew S. McCabe, Paul Cormican, Kate Keogh, Aaron O'Connor, Eoin O'Hara, Rafael Alejandro Palladino, David A. Kenny and Sinead M. Waters

Periodic feed restriction is used in cattle production to reduce feed costs, subsequently when normal feed levels are resumed; cattle typically display an accelerated growth phenomenon, known as compensatory growth (CG). Our objective was to examine microbial populations in rumen digesta following a period of dietary restriction and subsequent CG using Illumina Miseq Phylogenetic marker amplicon sequencing of DNA extracted from rumen contents. 60 Holstein Friesian bulls were fed either a restricted diet (125 days), following which they were fed *ad libitum*

for a further 55 days or fed *ad libitum* for the entirety of the trial. Rumen digesta samples were collected at the end of each period. Following a period of restricted feeding a large increase in overall species diversity, as well as an increase in the relative abundance of *Methanobrevibacter gottschalkii* clade were observed. A large decrease in the relative abundance of an uncharacterised Succinivibrionaceae species was also evident. The increase in the *Methanobrevibacter gottschalkii* clade was strongly correlated (ϱ=-0.72, $P<1x10^{-20}$) with the decrease in the Succinivibrionaceae species. Fifty five days of re-alimentation following previous dietary restriction, resulted in animals displaying CG, growing at 1.8 times that of their non-restricted counterparts. However although they displayed accelerated growth, their rumen microbes were not different to animals not restricted. This showed that the disruption caused by feed restriction did not have any long term effects on the rumen microbiome composition. Additionally, it also showed that the rumen prokaryotic microbiome was not associated with CG following 55 days of re-alimentation.

W492: Host-Microbe Interactions

Genome-Enabled Analysis of Pathogen Migration and Evolution

Erica M. Goss, University of Florida, Gainesville, FL

Genomics-based genetic marker discovery is leading to unprecedented collection of population genetic data for pathogens. Genomics-based data can be used for detailed analysis of the ancestry of population samples for studying pathogen molecular epidemiology and evolution. Population genomics studies of agricultural pathogens are increasing in number as the costs of sequencing continue to go down. Genomics-enabled markers are valuable for rapidly changing populations and for emerging and clonal pathogens that exhibit little genetic variation. International trade and continued globalization of agriculture has had major effects on plant pathogen populations and their interactions with crop hosts. Increased long-distance movement of pathogens has made disease management particularly challenging. Genomic information is providing data for disease management strategies. We are using whole genome sequencing to examine the processes underlying shifts in the *Xanthomonas perforans* population causing bacterial spot of tomato in Florida. We have found gain and loss of effectors not previously recognized as being under R gene selection, as well as variation in effector gene content correlated with strain genetic background, indicating multiple introductions of strains with different effector profiles. For understanding the role of strain movement in pathogen population shifts, national and international cooperative efforts are needed. Rapid monitoring of population variation using genomics-informed markers would benefit disease management in many pathosystems.

W493: Increasing Genetic Gains for Food Security in the Developing World

Developing Climate Resilient Crops and Livestock for the Developing World - Where are the Quantum Leaps?

Robert Bertram, US Agency for International Development, Washington, DC

Great gains have been made over the last 50 years in the struggle against poverty and hunger in the developing world, and substantial credit rests with international agricultural research – particularly crop improvement research. Now, despite great progress, nearly a billion people remain food insecure, and twice that number suffer from micronutrient deficiency. Moreover, the 21st century faces new challenges, including decreasing amounts of high-quality arable land, both per capita and in absolute terms, increased competition for limited water, and, most ominous due to its uncertain impacts, climate change. Smallholder farmers and herders will face more rapidly changing abiotic and biotic stresses than ever before. There is probably no more pressing opportunity for helping smallholder farmers to adapt to climate change than to increase the availability and accessibility of new crop varieties or livestock breeds with improved tolerance to heat, drought, flood, pests, and disease, either within a crop already grown, or of other, better adapted crops and livestock. While USAID and other investors work closely with national governments to build a supportive infrastructure to facilitate agriculture-led economic growth, such as transportation infrastructure and input supply systems, there is need for a focused research agenda to accelerate genetic gain in crops and livestock. How can we leapfrog the process of providing improved genetic materials at a pace that has not been possible before to mitigate the risks from climate change? Workshop participants are encouraged to reflect on what is needed to respond to these challenges for developing country agriculture, and consider what research is required to respond to those challenges.

W494: Increasing Genetic Gains for Food Security in the Developing World

Gates Foundation Investments and Objectives for Crop Improvement in the Developing World

Rob Horsch, Bill & Melinda Gates Foundation, Seattle, WA

W495: Increasing Genetic Gains for Food Security in the Developing World

How Can Advanced Genomics Change the Landscape for Breeding in the Developing World?

Edward S. Buckler, USDA-ARS-Cornell University, Ithaca, NY

Productive crops require that two things from genetics: (1) There genomes are free from deleterious mutations that have accumulated over time. (2) Their genetics is matched to local environment and agronomy. The tools of genomics, molecular biology, and population genetics are all giving quantitative genetics the information necessary to evaluate and breed crops with much greater efficiency. How will bioinformatics and modeling drive breeding where genomic selection and genome editing become common place? How to we ensure these benefits accrue for the developing world?

W496: Increasing Genetic Gains for Food Security in the Developing World

Accelerating Plant Breeding: Past, Present and Future

Jesse Poland, Kansas State University, Manhattan, KS

Since the start of modern plant breeding a century ago, there has been focused effort on how to optimize selections and maximize the rate of gain. Mid-century, during the heyday of quantitative genetics, advancement of the theory behind expected gain from selection was empirically explored and implemented to accelerate breeding. New statistical approaches to account for field variation lead to higher selection accuracy. During this time, the mechanization of breeding programs drastically increased population size and testing environments leading to more accurate assessments of breeding values and higher intensity of selection. With the strong foundation of robust field trials, the advent of

inexpensive molecular markers found fertile ground to push the speed and population sizes further. At the turn of this century, implementation of molecular breeding is currently enabling selection at a speed not possible with phenotypic selection. Spanning the scope from allele enrichment to whole-genome prediction, the breeding methodologies now possible through direct selection on genotypes gives new scope to breeding. Gazing into the future it is likely that we will see phenotyping once again coming front and center. Though, in this new century, phenotyping being automated and high-throughput at the intersection of classical breeding, physiology, engineering, machine vision, artificial intelligence and high-performance computing. Though genomics tools are revolutionizing breeding, in the end, the prediction models are only as good as the phenotypes that are fed in.

W497: Increasing Genetic Gains for Food Security in the Developing World

The Role of Genomics Tools for Genetic Improvement and Their Potential Application in Developing Countries

Dorian J. Garrick, Department of Animal Science, Iowa State University, Ames, IA

The performance of an agricultural industry in a particular management and environmental circumstance depends upon the suitability of the germplasm that comprised the founder nucleus population, ongoing genetic improvement in the nucleus since the founder generation was formed, and the genetic lag that exists between the nucleus and the commercial populations. In developed countries, ongoing genetic improvement in most species relies on multiple trait selection, traditionally based on measured phenotypes of large cohorts of individuals within a pedigree structure, but these analyses now commonly enhanced using whole-genome SNP information. Genetic lag is minimized by exploiting reproductive technologies such as artificial insemination and/or embryo transfer. In developing countries, pedigree information can be problematic to collect, cohorts of contemporaries are often of inadequate size, and advanced reproductive technologies can be difficult to employ. In developing countries, case-control protocols for use in genome-wide association studies offer promise for identifying genomic regions with major gene effects that could be used for marker-assisted selection or introgression. In both developed and developing countries, genomic tools such as exome annotation offer promise for identifying mutations that might be useful to introduce into nucleus populations by introgression or preferably by gene editing.

W498: Increasing Genetic Gains for Food Security in the Developing World

Increasing the Rate of Genetic Gain Delivered to Farmers: Some Lessons from US Commercial Plant

Roy G. Cantrell, Monsanto, Chesterfield, MO

The focus of Monsanto Breeding is to increase genetic gain and protect productivity of six crops globally. This is built on a base of extensive genetic diversity that can be harnessed with new genomic and data technologies. These integrated systems enabled accelerated improvement in key plant health traits. The application of Genome Wide Selection permits scaling of breeding programs. Layers of data require interconnected analytics to enhance decision making in breeding programs. The extensive data collected on products in the pipeline not only support better product advancement but provides valuable data on products released to support decisions by the customer. Commercial plant breeding continues to undergo transformation that will overtime have a global impact across multiple crops.

W499: Increasing Genetic Gains for Food Security in the Developing World

Key Success Factors and Challenges on How to Integrate Genotyping and Modern Phenotyping Tools to Improve the Efficiency, Accelerate, and Lower the Cost of a Breeding Program in Haiti

Gael Pressoir, Fondation CHIBAS, , Bon Repos Croix des Bouquets, Haiti

Chibas is an institute involved in breeding for the subhumid and dryland tropics and more specifically for Haiti. Chibas has ongoing programs breeding multipurpose sweet sorghum (grain, stem sugars and forage), edible Jatropha (oil and high protein meal as a soymeal substitute), and has just initiated new programs in common beans, peanuts and perennial pigeon pea.

Working on orphan perennial crops such as edible Jatropha and perennial pigeon pea, Chibas is increasingly using novel technologies such as genotyping by sequencing and modern phenotyping tools such as NIR to shorten the duration of our selection cycles and to reduce the cost of breeding. We will showcase how these technologies have allowed us in just five years to (1) study Jatropha curcas genetic diversity and to simultaneously (2) develop a working male sterility system, (3) introgress the male sterility and edible alleles in different populations, and (4) also develop heterotic genepools to carry out a reciprocal recurrent selection scheme. In adopting these technologies the limiting factor is not so much the cost of these technologies (we actually save time and money); it has more to do with the lack of adequate human resources, working in a cash deprived research environment and the lack of funding for international collaborative research program that allow this transfer of, and access to, technology (access to a genotyping by sequencing platform for example). Challenges and solutions to using these technologies in a breeding program in Haiti will be presented.

W500: Increasing Genetic Gains for Food Security in the Developing World

Panel Discussion: Successes and Challenges in Applying Genomics Tools in Developing Country Breeding Programs

Mitch Tuinstra, Purdue University, West Lafayette, IN

W501: Increasing Genetic Gains for Food Security in the Developing World

Panel discussant

Ndiaga Cissé, ISRA-CERAAS, Thies, Senegal

W502: Increasing Genetic Gains for Food Security in the Developing World

Panel discussant

Clare T. M. Mukankusi, International Centre for Tropical Agriculture (CIAT)-Uganda, Kampala, Uganda

W503: Increasing Genetic Gains for Food Security in the Developing World

Panel discussant
Paul Gibson, Makerere University, Kampala, Uganda

W504: Increasing Genetic Gains for Food Security in the Developing World
Panel discussant
Roy G. Cantrell, Monsanto, Chesterfield, MO

W505: Increasing Genetic Gains for Food Security in the Developing World
Panel discussant
Heather Burrow, Cooperative Research Centre for Beef Genetic Technologies, Armidale, Australia

W506: Increasing Genetic Gains for Food Security in the Developing World
Tools and Services for the Application of Genomics to Cultivar Development in Africa and South Asia
Kelly Robbins, Cornell University, Ithaca, NY

W507: Increasing Genetic Gains for Food Security in the Developing World
Increasing the Rate of Genetic Gain in the CIMMYT Maize Program Using Doubled Haploids and Marker-Assisted Selection.
Michael S. Olsen, CIMMYT, Nairobi, Kenya
The CIMMYT Global Maize Program (GMP) is applying doubled haploid technology and marker-assisted breeding approaches both jointly and independently to increase the rate of genetic gain for biotic and abiotic stress tolerance in our breeding programs. Key considerations regarding best practices for successful implementation of these technologies in applied breeding programs will be explored. Emphasis will be on logistics planning, critical path analysis, decision tree flowcharts, and stakeholder engagement for optimizing the choice of method or combination of methods to maximize the cost effectiveness of the breeding programs. A case study looking at the integrated application of these technologies to the GMP-Africa breeding programs will be presented. The challenge of the GMP-Africa team is to combine multiple biotic stress tolerances, particularly Maize Streak Virus (MSV) and Maize Lethal Necrosis (MLN), together with appropriate abiotic stress tolerances, competitive yield under optimal conditions, and improved cost of goods sold (COGS) parameters. The breeding team utilizes a combination of elite adapted and relatively unadapted donor germplasm (both temperate and tropical) to simultaneously improve genetic gain for each of these important traits.

W508: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
Update on the Development and Deployment of the Breeding Management System
Jean-Marcel Ribaut, Integrated Breeding Platform, Texcoco, Mexico
The Integrated Breeding Platform (IBP) – www.integratedbreeding.net – is a comprehensive source for best practices in plant breeding, where registered users can find tools, services, capacity development and community resources. At the core of this offer is the IBP Breeding Management System (BMS), a software suite of integrated applications that facilitates the conduct of typical, modern plant breeding activities. The BMS contains a database that manages data associated with plant breeding, including pedigree information, seed inventory, phenotype and genotype data, and comes preloaded with trait, location and breeding method ontologies. Over the last year, changes have been brought to the architecture of the system for the development of a web-based version (BMS v4). Thanks to improved concurrency, translation and system performance, the BMS can now comfortably support 10-20+ concurrent users, as well as trials of up to 500 entries in 200 sites. Feature improvements also include an improved seed inventory system, support for an expanded set of handheld devices, import of custom field designs, a better ontology management system, and more efficient historical data management capabilities. Building on the robust stand-alone version released in February 2015 (v3.0.8), and on the upcoming public release of the web version (v4), the IBP team and its Regional Hubs are now actively deploying the BMS in various national breeding programs. It is also being deployed at several CGIAR Centers including CIMMYT, ICRISAT, IITA and Africa Rice. The BMS is accessible to both public and private breeding programs.

W509: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding
Modernizing Public Plant Breeding Programs to Deliver Higher Rates of Genetic Gain to Farmers in the Developing World
Gary Atlin, Bill & Melinda Gates Foundation, Seattle, WA
New breeding technologies and management approaches have revolutionized plant breeding in commercial cropping systems in temperate countries, but these have not been widely applied in public plant breeding programs in the developing world. These programs will provide smallholder farmers in Africa, Asia, and Latin America with seed of the self-pollinated, open-pollinated, and vegetatively-propagated crops on which they depend for many years to come. Rates of genetic gain generated by breeding programs and delivered to farmers, although rarely measured, appear to be very low. Most public plant breeding programs have cycle times that are too long, are not optimized with respect to advancement rates and allocation of testing resources, do not have access to an integrated breeding informatics database, make no use of DNA markers in forward breeding, lack formal product concepts, and do not generate enough high-quality data to confidently recommend and promote their products as being clearly superior to varieties in use. It is critical that public sector breeding programs modernize both their breeding technology and their management methods if rates of genetic gain high enough to feed still-growing populations in the face of climate change are to be achieved. The key metrics by which breeding programs and seed systems in developing countries need to be assessed are (i) the rate of genetic gain they deliver in farmers' fields, and (ii) the average age of varieties in farmers' fields. The Gates Foundation is supporting several initiatives to provide support on breeding strategy, breeding informatics tools, low cost genotyping, pipeline optimization, and dissemination planning to CGIAR and national breeding programs in generating and delivering higher rates of genetic gain for smallholder farmers in Africa and South Asia.

W510: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding

The Crop Ontology: A Source of Standard Traits and Variables for Breeding and Agronomy

Elizabeth Arnaud[1], Léo Valette[1], Julian Pietragalla[2], Marie-Angélique Laporte[1], Céline Aubert[1], Medha Devare[3], Graham McLaren[2] and Jean-Marcel Ribaut[2], (1)Bioversity International, Montpellier Cedex 5, France, (2)Integrated Breeding Platform, Texcoco, Mexico, (3)CGIAR Consortium Office, Montpellier, France

The Crop Ontology is a service of the Integrated Breeding Platform (www.integratedbreeding.net) in collaboration with the CGIAR and partners and under the leadership of Bioversity international. The Crop Ontology (www.cropontology.org) provides harmonized and validated breeders' trait names, measurement methods, scales for currently 18 crops that are used by the Breeding Management System (BMS). The NextGeneration Breeding Databases developed by Boyce Thompson Institute also embed the Crop Ontology traits. The Crop Ontology contributes to the content of the reference ontologies of the Planteome project (http://www.planteome.org/).

A new Trait Dictionary Template was released that now includes the '**standard variable**'. A standard variable is equal to '*one trait+one method+one variable*' and a trait can be measured through different variables, according to the method or the scale used. These variables will accurately annotate the measurements stored in the BMS databases and also will support the creation of standard manual or electronic fieldbooks. Ten crop trait dictionaries have already been migrated into this new template and uploaded on the Crop ontology site. Using similar methodology, an Agronomy Ontology is being developed to support combining results of field management practices with crop traits which is important to fully understand the dynamic of varying factors within any cropping system. Curation is currently performed to secure the compliance between the Agronomy ontology and the variables of the International Consortium for Agricultural Systems Applications (ICASA).

W511: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding

BRAPI, a Community-Based Effort for Standardizing Breeding Application Interfaces

Lukas Mueller, Boyce Thompson Institute for Plant Research, Ithaca, NY

As genome-based methods for breeding have been established over the last decade, breeding data storage needs have soared. Comprising sequence, marker, genotyping calls, phenotypic information, and trial metadata, the breeding data are relatively complex and potentially extremely voluminous. Most breeding data have been stored in custom database systems that have been created for different breeding programs and centers, often developed independently with little attention to compatibility with other systems. To leverage the power of these databases, we have organized the BRAPI consortium, consisting of major providers of public breeding database solutions as well as their users. The goal of BRAPI to create web-based application program interface (API) that can be used by breeder software providers as a standardized interface to breeding data, independent of the data source, making the creation of tools and the exchange of breeding data easier, as well as building a more tightly linked plant breeding software community.

W512: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding

Breeding Data Management & Experience of BMS deployment at ICRISAT

Abhishek Rathore[1], Praveen Reddy T.[1], Sravani Mana[1], Sarita Pandey[1], Vikas K Singh[1], Manish Roorkiwal[1], Santosh Deshpande[1], Anu Chitikineni[1], BhanuPrakash Amindala[1], Mohan Telluri[1], Srinivasarao Chukka[1], Arllet Portugal[2], Stefania Grando[1] and Rajeev K Varshney[1], (1)ICRISAT, Hyderabad, India, (2)Integrated Breeding Platform, Texcoco, Mexico

Research data management stands for use of appropriate experimental design and proper capturing, analysis, interpretation, publishing, archiving and sharing of data. Data management is critical for research organizations to ensure research quality and integrity. For public plant breeding programs proper data management has been a key challenge for various reasons and one of them is non-availability of public data management software platforms. Commercially available software are often expensive and non-affordable at NARS of developing countries. Integrated Breeding Platform (IBP) addressed this challenge by providing the state of art suite of software called Breeding Management System (BMS). It has several features including recording of pedigrees and tracking of genealogies, cross generation and advancing nurseries, generating field layouts and statistical analysis. BMS has been made available for public and commercial plant breeding organizations and can be installed either standalone or on enterprise cloud. Because of these and several other features, ICRISAT has decided to migrate breeding program data to BMS and adopt the software as the breeding management platform. We are establishing central cloud based BMS repository for the institute including multi-institutional projects such as TL-III and HOPE 2. This migration involves inventorying historical data, uploading pedigrees, digital data capture and barcoding. ICRISAT is developing APIs to link BMS to data platforms like Dataverse and a high throughput genotyping data management system under GOBII project. In this context, we have already started data migration at ICRISAT Headquarter in early 2015 and gradually rolling out to data from our locations in sub-Saharan Africa.

W513: Integrated Breeding Platform: Tools, Databases and Applications for Plant Breeding

Plant Breeding E-Learning in Africa – Developing and Delivering High Quality E-Learning Resources That Match Local Needs

Walter P. Suza, Iowa State University, Ames, IA

A key element of increasing agricultural productivity, enhancing food security, and improving rural incomes in Africa is the more rapid and effective delivery of improved, farmer-preferred varieties. Practical plant breeders who can independently manage breeding programs with a good understanding of the seed business are central to that process. However, the number of such breeders is limited in sub-Saharan Africa (SSA). Therefore, there is a need for increased training of plant breeders, especially at the MSc level, who will play a key role in contributing toward food security in Africa. In 2011 the Agronomy Department at Iowa State University began to offer an online MS degree in Plant Breeding that is modeled after the highly successful MS degree in Agronomy which began in 1998. Plant Breeding E-Learning in Africa (PBEA) was created to develop and deliver e-learning modules containing relevant learning features, graphics, videos, and applied learning activities to contribute to the training of plant breeders in Africa. Phase 1 of PBEA has focused on adapting courses from ISU's online MS in Plant Breeding in partnership with universities in SSA and Integrated Breeding Platform (IBP). In addition to topics on crop genetics, quantitative genetics, molecular plant breeding, crop improvement, and quantitative methods, a module on teaching of IBP's Breeding Management System has been

developed for distribution in 2016. Feedback from the African faculty and students indicates PBEA e-learning resources are of great value and substantially relevant to teaching needs in Africa.

W514: International Cotton Genome Initiative;(ICGI)

Introduction and Update

John Z. Yu, USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX and David M. Stelly, Texas A&M University, College Station, TX

Update to ICGI

W515: International Cotton Genome Initiative;(ICGI)

Comparative Genomic De-Convolution of the Cotton Genome Revealed a Decaploid Ancestor and Widespread Chromosomal Fractionation

Xiyin Wang, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

The 'apparently' simple genomes of many angiosperms mask complex evolutionary histories. The reference genome sequence for cotton (Gossypium spp.) revealed a ploidy change of a complexity unprecedented to date, indeed that could not be distinguished as to its exact dosage. Herein, by developing several comparative, computational, and statistical approaches, we revealed a 5x multiplication in the cotton lineage of an ancestral genome common to cotton and cacao and proposed evolutionary models to show how such a decaploid ancestor formed. The ~70% gene loss necessary to bring the ancestral decaploid to its current gene count appears to fit an approximate geometrical model, that is, though many genes may be lost by single-gene deletion events, some may be lost in groups of consecutive genes. Gene loss following cotton decaploidy has largely just reduced gene copy numbers of some homologous groups. We designed a novel approach to deconvolute layers of chromosome homology, providing definitive information on gene orthology and paralogy across broad evolutionary distances, both of fundamental value and serving as an important platform to support further studies in and beyond cotton and genomics communities.

W516: International Cotton Genome Initiative;(ICGI)

TAL Effector-Mediated Susceptibility to Bacterial Blight of Cotton

Kevin Cox Jr.[1], Fanhong Meng[1], Katie Wilkins[2], John Z. Yu[3], Ping He[1], Adam Bogdanove[2] and Libo Shan[1], (1)Texas A&M University, College Station, TX, (2)Cornell University, Ithaca, NY, (3)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX

Bacterial blight of cotton (BBC) caused by *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) is a destructive disease that has recently re-emerged in the U.S. *Xcm* injects transcription activator-like (TAL) effectors that directly induce the expression of host susceptibility (*S*) or resistance (*R*) genes. Although more than ten *Xcm* TAL effectors have been characterized, no corresponding *S* or *R* genes have been identified in cotton. Avrb6, a key TAL effector in *Xcm*, plays a major role in promoting pathogenicity in cotton susceptible lines. By using a systems approach combining genome-wide transcriptome profiling and TAL effector binding site (EBE) computational prediction, we found that *GhSWEET1*, encoding a sugar transporter, is highly induced in cotton by *Xcm* Avrb6. Importantly, the promoter of *GhSWEET1* possesses an EBE site that is specifically activated by Avrb6. Silencing of *GhSWEET1* in cotton via virus-induced gene silencing (VIGS) showed reduced susceptibility to the infections by *Xcm* carrying *avrb6*, whereas activation of *GhSWEET1* by designed TAL effectors enhanced susceptibility of *Xcm*. Thus, *GhSWEET1* is an *S* gene of *Xcm avrb6* for promoting *Xcm* pathogenicity in cotton. GhSWEET1 contributes to pathogenicity likely via sugar efflux to apoplast to promote *Xcm* proliferation. We are deploying CRISPR-CAS to engineer the EBE of GhSWEET1 to avoid the induction by Xcm, thereby biologically controlling this important disease.

W517: International Cotton Genome Initiative;(ICGI)

Combination of Next Generation Mapping, Transcriptome and Functional Analyses: A New Strategy to Identify Candidate Genes for Fiber Quality Traits in Upland Cotton

Md Sariful Islam[1], David D. Fang[1], Linghe Zeng[2], Gregory Thyssen[3], Chris D. Delhom[4] and Hee Jin Kim[1], (1)Cotton Fiber Biosciences Unit, USDA-ARS-SRRC, New Orleans, LA, (2)USDA ARS, Stoneville, MS, (3)Cotton Chemistry and Utilization Unit, USDA-ARS-SRRC, New Orleans, LA, (4)USDA ARS SRRC, New Orleans, LA

Fiber strength, length, maturity and fineness determine the market value of cotton fibers and the quality of spun yarn. Cotton fiber strength has been recognized as a critical quality attribute in the modern textile industry. Molecular mechanisms responsible for regulating fiber strength still remain unclear. *Gossypium hirsutum* near isogenic lines (NILs), MD52ne and MD90ne showing variations in fiber strength provide an opportunity for uncovering the molecular and genetic basis of superior fiber quality. Comparative transcriptome analyses of the NILs showed that the superior bundle strength of MD52ne fibers was potentially related to two signaling pathways: one is ethylene and the interconnected phytohormonal pathways that are involved in cotton fiber elongation, and the other is receptor-like kinases (RLKs) signaling pathways that are involved in maintaining cell wall integrity. A group of 27 new SNP markers generated from mapping-by-sequencing (MBS) were placed in QTL regions to improve existing maps. The QTL regions contained multiple significantly differentially expressed RLKs. SNPs that result in non-synonymous substitutions to amino acid sequences of annotated genes were identified within these differentially expressed genes, and mapped. Taken together, mapping, transcriptome and amino acid mutation analysis indicate that RLK genes that were suggested to mediate a coordination of cell elongation and SCW biosynthesis in other plants likely are candidate genes for regulating cotton fiber cell wall assembly and strength. MBS along with transcriptome analysis demonstrated a powerful strategy to elucidate candidate genes for the QTLs that control complex traits in a complex genome like tetraploid upland cotton.

W518: International Cotton Genome Initiative;(ICGI)

Comparative Analysis of Genome-Wide Divergence, Domestication Footprints and Genome-Wide Association Study of Root Traits for *Gossypium hirsutum* and *Gossypium barbadense*

Umesh K. Reddy[1], Padma Nimmakayala[1], Lakshmi V. Abburi[1], CVC Mohan Reddy[1], Thangasamy Saminathan[1], Richard G. Percy[2], John Z. Yu[2], James Frelichowski[2], Josh Udall[3], Justin T. Page[3], Dong Zhang[4], Tariq Shehzad[4] and Andrew H. Paterson[4], (1)Department of Biology, West Virginia State University, Institute, WV, (2)USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX, (3)Brigham Young University, Provo, UT, (4)Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

Use of 10,129 singleton SNPs of known genomic location in tetraploid cotton provided unique opportunities to characterize genome-wide diversity among 440 *Gossypium hirsutum* and 219 *G. barbadense* cultivars and landrace accessions of widespread origin. Using the SNPs distributed genome-wide, we examined levels and patterns of genetic diversity, haplotype distribution and runs of homozygosity (ROH) in the *G. hirsutum* and *G. barbadense* genomes to clarify the genomic architecture, domestication process and population demographic history. Diversity and identity-by-state analyses revealed little sharing of alleles between the two cultivated allotetraploid genomes, with a few exceptions which indicated sporadic gene flow. An important finding was the location of a high number of selective sweeps, as represented by low diversity regions common to both species, in D-genome–derived chromosomes. The presence of conserved linkage disequilibrium (LD) blocks, haplotypes and ROHs between *G. hirsutum* and *G. barbadense* provide strong evidence for comparable patterns of evolution, either in their domestication processes or perhaps during the radiation of tetraploid species. Our comparative genome-wide association study (GWAS) of a set of seedling root traits revealed common SNPs in *G. hirsutum* and *G. barbadense*, which strengthens the evidence for convergent evolution. Our study illustrates the potential use of population genetic techniques to identify genomic regions for cotton breeding.

W519: International Cotton Genome Initiative;(ICGI)

High-Quality Draft Genome of *Gossypium herbaceum* cv. Wagad

Thiruvarangan Ramaraj[1], Aaron Sharp[2], Joann Mudge[1] and Josh Udall[2], (1)National Center for Genome Resources (NCGR), Santa Fe, NM, (2)Brigham Young University, Provo, UT

As part of a larger effort to investigate structural variation in assorted diploid and polyploidy cotton genomes we have sequenced and assembled *Gossypium herbaceum*. Cultivated *Gossypium herbaceum* is a A-genome diploid from the Old World with a genome size of approximately 1.7 Gbp. Here we present a high quality draft genome of *Gossypium herbaceum* (cv. Wagad) using a multi-platform sequencing strategy (Illumina HiSeq, Dovetail Genomics, PacBio RS II, and BioNano Whole Genome Maps). The multi-platform sequencing strategy will help immensely in achieving high quality *de novo* reconstructions of genomes, which requires accurate long-range contiguity. The sequence data and the final assembly will be used towards comparative analysis with *Gossypium arboreum*, which is also a domesticated A-genome diploid. The final goals of this project is to provide a high quality reference genome for *Gossypium herbaceum*, assess recent technologies such as Dovetail Genomics and BioNano Genomics and also, serve as a model to the plant genomics community who has an interest in using multi-platform sequencing technologies for *de novo* genome sequencing.

W520: International Cotton Genome Initiative;(ICGI)

1-Minute Oral Presentations on Cotton Genomics

John Z. Yu, USDA-ARS, Southern Plains Agricultural Research Center, College Station, TX and David M. Stelly, Texas A&M University, College Station, TX

This 20-minute time slot will feature a series of 15 fast-paced oral introductions of cotton genomics-relevant PAG posters, by the presenter of the respective poster. Each presentation will be strictly oral, and have the duration of about one minute to highlight key points of the individual poster presentation on cotton genomics. Besides learning about each of the posters, attendees will be better able to subsequently recognize presenters, and know a little bit more about their cotton genomics research.

W521: International Goat Genome Consortium

Introduction

Gwenola Tosser-Klopp, INRA, Castanet-Tolosan, France

W522: International Goat Genome Consortium

From Sequencing to Chromosomes: New *de novo* Assembly and Scaffolding Methods Improve the Goat Reference Genome

Sergey Koren[1], Derek M. Bickhart[2], Adam M Phillippy[1], Timothy P.L. Smith[3], Shawn T. Sullivan[4], Ivan Liachko[5], Joshua N. Burton[5], Maitreya J. Dunham[5], Jay Shendure[5], Alex R. Hastie[6], Brian L. Sayre[7], Heather J Huson[8], George E. Liu[2], Benjamin D. Rosen[2], Steven G. Schroeder[2], Curtis P. VanTassell[2] and Tad S. Sonstegard[9], (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)USDA, ARS, USMARC, Clay Center, NE, (4)Phase Genomics, Seattle, WA, (5)University of Washington - Department of Genome Sciences, Seattle, WA, (6)BioNano Genomics, San Diego, CA, (7)Virginia State University, Petersburg, VA, (8)Cornell University, Ithaca, NY, (9)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN

Single-molecule sequencing is now routinely used to assemble complete, high-quality microbial genomes, but these assembly methods have not scaled well to large genomes. To address this problem, we previously introduced the MinHash Alignment Process (MHAP) for overlapping single-molecule reads using probabilistic, locality-sensitive hashing. Integrating MHAP with Celera Assembler (CA) has enabled reference-grade assemblies of model organisms, revealing novel heterochromatic sequences and filling low-complexity gap sequences in the GRCh38 human reference genome. We have applied our methods to assemble the San Clemente goat genome. Combining single-molecule sequencing from Pacific Biosciences and BioNano Genomics generates and assembly that is over 150-fold more contiguous than the latest *Capra hircus* reference. In combination with Hi-C sequencing, the assembly surpasses reference assemblies, *de novo*, with minimal manual intervention. The autosomes are each assembled into a single scaffold. Our assembly provides a more complete gene reconstruction, better alignments with Goat 52k chip, and improved allosome reconstruction. In addition to providing increased continuity of sequence, our assembly achieves a higher

BUSCO completion score (84%) than the existing goat reference assembly suggesting better quality annotation of gene models. Our results demonstrate that single-molecule sequencing can produce near-complete eukaryotic genomes at modest cost and minimal manual effort.

W523: International Goat Genome Consortium

How to Sequence an Ancient Goat Genome

Kevin Daly, Trintiy College Dublin, Dublin, Ireland

Goat domestication from wild bezoar is thought to have occurred approximately 10,000 years ago in the highlands of Iran and Turkey. This process has had a profound effect on this species‘ production traits, behaviour, and physical appearance, which will have left distinct genetic signals in the modern domestic population. However, such signals will have been confounded by human-driven migration, post-domestication admixture, and perhaps multiple domestication events. Directly sequencing the DNA of ancient goats, preserved in bone remains, should allow more detailed unpacking of these different processes. Postmortem DNA fragments and degrades in an environment-dependent manner and the locations of key caprine population events are challenging in this respect. However, recent studies have demonstrated that denser bone elements can provide a refuge from exogenous sources of damage for DNA. By extracting DNA from such bones, we will show it is possible to generate entire genomes from geographic locations previously thought to be refractory to DNA preservation.

W524: International Goat Genome Consortium

The Five Goats of Eve: The Impact of Domestication on the Goat Mitogenome Variability

Licia Colli[1,2], Hovirag Lancioni[3], Irene Cardinali[3], Anna Olivieri[4], Marco Rosario Capodiferro[4], Marco Pellecchia[1], Marcin Rzepus[1,5], Wahid Zamani[6,7], Saeid Naderi[8], Francesca Gandini[4,9], Seyed Mohammad Farhad Vahidi[10], Saif Agha[11], Ettore Randi[12,13], Vincenza Battaglia[4], Maria Teresa Sardina[14], Baldassare Portolano[14], Hamid Reza Rezaei[15], Petros Lymberakis[16], Frédéric Boyer[6], Eric Coissac[6], François Pompanon[6], Pierre Taberlet[6], Paolo Ajmone Marsan[1,2] and Alessandro Achilli[3,4], (1)Inst. of Zootechnics, Università Cattolica del S. Cuore, Piacenza, Italy, (2)Research Center on Biodiversity and Ancient DNA – BioDNA, Piacenza, Italy, (3)Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, Perugia, Italy, (4)Dipartimento di Biologia e Biotecnologie “L. Spallanzani”, Università di Pavia, Pavia, Italy, (5)Institute of Food Science and Nutrition - ISAN, Università Cattolica del S. Cuore, Piacenza, Italy, (6)LECA - CNRS - Université de Grenoble 1, Grenoble, France, (7)Department of Environmental Sciences, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Mazandaran, Iran, (8)Natural Resources Faculty, University of Guilan, Guilan, Iran, (9)School of Applied Sciences, University of Huddersfield, Huddersfield, United Kingdom, (10)Agricultural Biotechnology Research Institute of Iran (ABRII), Rasht, Iran, (11)Department of Animal Production, Faculty of Agriculture, Ain Shams University, Cairo, Egypt, (12)Laboratorio di Genetica, Istituto per la Protezione e la Ricerca Ambientale (ISPRA), Ozzano dell'Emilia (BO), Italy, (13)Department 18/Section of Environmental Engineering, Aalborg University, Aalborg, Denmark, (14)Dipartimento Scienze Agrarie e Forestali, Università degli Studi di Palermo, Palermo, Italy, (15)Environmental Sciences Department, Gorgan University of Agriculture and Natural Resources, Gorgan, Iraq, (16)Natural History Museum of Crete, University of Crete, Iraklio, Crete, Greece

Domestic goat (*Capra hircus*) is among the most important and diffused livestock species, whose mitochondrial gene pool was dramatically shaped during the first phases of domestication. So far, the analysis of goat control region sequences identified 6 different haplogroups and highlighted a lower geographic structuring of diversity compared to other livestock species. Here, we present the first extensive survey of goat mitochondrial variability based on 84 complete mitogenomes of 79 *C. hircus* and 5 *Capra aegagrus* individuals. Phylogenetic analyses dated the most recent common ancestor to ~460 kya and revealed five distinctive domestic haplogroups (A, B1, C1a, D1 and G), clearly nested within wild goat branches. These five clades almost simultaneously diverged at the interface between the Epipaleolithic and early Neolithic periods and dramatically expanded ~12-10 ka ago. Thus, similarly to cattle and other livestock, it is conceivable that also in goat only few female founder lineages underwent domestication after surviving the last glacial maximum in the Near Eastern *refugia*. Zooarchaelogical and ancient DNA data indicate Southeastern Anatolia as the most likely center of this primary domestication and confirm that the *C. hircus* populations brought by the first Neolithic migration waves into the Mediterranean were already characterized by at least two ancestral A and C variants. Actually, the ancient phylogenetic separation of C branch (~130 ka ago) points to a genetically distinct population that could have been involved in a secondary domestication event in Western Iran where haplogroup C is still common among bezoars.

W525: International Goat Genome Consortium

High Density SNP Chips Array Uncovers Genetic Diversity and Population Structure of 16 Ethiopian and Chinese Goat Populations

Getinet Mekuriaw[1], Joram Mwacharo[2], Kassahun Tesfaye[1], Dessie Tadelle[3], Mwai Okeyo[4], Appolinaire Djikeng[5], Bin Liu[6], Sarah Osama[7], Christine Grossen[8] and Wenguang Zhang[6,9], (1)Addis Ababa University, Addis Ababa, Ethiopia, (2)International Centre for Agricultural Research in the Dry Areas, Addis Ababa, Ethiopia, (3)International Livestock Research Institute, Addis Ababa, Ethiopia, (4)International Livestock Research Institute, Nairobi, Kenya, (5)BecA-ILRI Hub, Nairobi, Kenya, (6)Inner Mongolia Agricultural University, Hohhot, China, (7)Bioscience for eastern and central Africa, Nairobi, Kenya, (8)Institute of Evolutionary Biology and Environmental Studies, Zürich, Switzerland, (9)International Goat Genome Consortium, Hohhot, China

Genome wide survey was conducted to assess genetic diversity and population structure of 16 goat populations from Ethiopia and China. A total of 468 samples, which were genotyped using GoatSNP50k BeadCHIP panel were included in the study. Moreover, from the wild goat, Swiss alpine Ibex was also included as the 17th study population for comparison. However, due to the highest monomorphic (99.6%) loci accumulation, the Swiss alpine Ibex was limited only for the SNP polymorphism and diversity study. The average Ho and H_E were 0.375±0.1 and 0.383±0.00, respectively. The level of inbreeding ranged from -0.020 (Ibex-Cashmere hybrid) to 0.073 (Nubian goat). Average monomorphic and polymorphic loci were estimated to be 1749.19±627.37(3.5%) and 48000±627.52(96.5%). Molecular variance revealed 11.92% of variation among Chinese and Ethiopian goat populations studied. However, low (≤1%) population differentiation, lowest F_{ST}(μ=0.026) and highest *N*m (μ=24.76) were observed among Ethiopian goats. The PCA1&2 differentiated Ethiopian goats, Cashmere and Ibex hybrid. Similarly, PCA1&3

differentiated Kaffa goat from other Ethiopian goats. The ADMIXTURE analysis revealed presence of only six genetic backgrounds in Ethiopian goats. Overall, high genetic diversity and weak population structure were observed in Ethiopian goat populations. The gene flow facilitated through animal exchange could be one of the possible reasons for the observed pattern.

W526: International Goat Genome Consortium

African Goat Improvement Network: Community-Based Breeding Programs for Sustainable Genetic Improvement

Benjamin D. Rosen[1], Heather J Huson[2], Elizabeth A. Staiger[2], Tad S. Sonstegard[3], Jeffrey T. Silverstein[4], Brian L. Sayre[5], M.J. Woodward-Greene[6], Steven G. Schroeder[7], Gordon Spangler[7], Erin E. Connor[7], Timothy Gondwe[8], Max F. Rothschild[9], Henry Aaron Mulindwa[10], Kahsa Tadel Gebre[11], Khanyisile Mdladla[12], Tadele Mirkena[13], Farai C. Muchadeyi[14], Johann Soelkner[15] and Curtis P. VanTassell[7], (1)ARS, USDA, Beltsville, MD, (2)Cornell University, Ithaca, NY, (3)Acceligen Inc. Animal Ag. Subsidary of Recombinetics, St. Paul, MN, (4)USDA-Agricultural Research Service, Beltsville, MD, (5)Virginia State University, Petersburg, VA, (6)USDA-ARS, Beltsville, MD, (7)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (8)Lilongwe University of Agriculture and Natural Resources, Lilongwe, Malawi, (9)Iowa State University, Ames, IA, (10)National Livestock Resources Research Institute, Uganda, Tororo, Uganda, (11)Mekelle University, Mekelle, Ethiopia, (12)University of KwaZulu-Natal, Pietermartzburg, South Africa, (13)Hawassa University, Hawassa, Ethiopia, (14)Agricultural Research Council, Pretoria, South Africa, (15)University of Natural Resources and Life Sciences, Vienna, Austria

Goats are a critical component of smallholder farming in Africa. Their small size makes them relatively easy to acquire and maintain while their browsing habit means they are particularly well suited to the marginal crop regions of sub-Saharan Africa. Indeed, goats have the fastest growth rate of all livestock species in many regions. However, indigenous goats, while adapted to local climate and disease pressures, are not nearly as productive as their European counterparts. Past efforts to improve productivity have focused on a strategy to replace or crossbreed indigenous goats with imported breeds. Research has shown that this type of intervention may offer short term gains but frequently fail in the long run. Community-based breeding programs (CBBPs) may be a better solution. CBBPs organized around the principle of improving indigenous genetics through breeding and selection have been proven to be the best way to obtain permanent gains. To this end, we are establishing a total of four CBBPs in Uganda and Malawi through the National Agricultural Research Organization (NARO) and Lilongwe University of Agriculture and Natural Resources (LUANAR) with collaboration of the University of Natural Resources and Life Sciences, Vienna (BOKU). In preparation for developing locally individualized breeding strategies we are collecting genotype and phenotype data from different goat breeds across Africa. Our initial results suggest that there are very strong geographic influences to population structure, and that inbreeding may be an issue to contend with in certain breeding populations.

W527: International Goat Genome Consortium

Genome-Wide Association Study of Conformation and Milk Yield in Mixed-Breed Dairy Goats.

Sebastian L. Mucha, Raphael Mrode, Mike Coffey and Joanne C. Conington, Scotland's Rural College, Easter Bush, United Kingdom

Identification of genetic markers that affect economically important traits is of high value both from a biological point of view (identification of candidate genes) as well as practical benefits for the industry (genomic selection). This study is one of the first to investigate genetic background of economically important traits in dairy goats using the caprine 50k SNP chip. The aim of the project was to perform a genome-wide association study for milk yield and conformation of udder, teat, legs and feet. A total of 4563 goats had conformation and milk yield data, out of which 402 were genotyped with the Illumina Caprine 50K BeadChip. GWAS was performed using the Multi-Locus Mixed Model (MLMM) algorithm implemented in SNP & Variation Suite v7.7.8. Genome-wise significant SNP for milk yield was identified on chromosome 19, with additional chromosome-wise significant SNP on chromosomes 4, 8, 14, 28, 29 . Three genome-wise significant SNP for conformation were identified on chromosome 19, and chromosome-wise SNPs were found on chromosomes 4, 5, 6, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 23, 27, 28. The proportion of variance explained by the significant SNPs was between 0.4 to 15.0%. This study is the first attempt to identify SNP associated with milk yield and conformation in dairy goats. Due to small sample size, associations with candidate genes should be treated as an indication and certainly require further research to be validated.

W528: International Goat Genome Consortium

Design of a SNP Parentage Assignment Panel for French Goat Breeds

Isabelle Palhière, INRA-GenPhySE, Toulouse, France, Flavie Tortereau, INRA, Castanet Tolosan Cedex, France, Pierre Martin, Capgenes, MIGNALOUX-BEAUVOIR, France, Rachel Rupp, INRA GENPHYSE, Toulouse, France and **Gwenola Tosser-Klopp**, INRA, Castanet-Tolosan, France

Knowing pedigrees is essential for selection, conservation and management of animal populations. In French goats, complete and accurate pedigree information is difficult to get because of the increasing use of multiple-sired matings. In the two main French breeds (Alpine and Saanen), only 50% of females involved in the official genetic evaluation have a known sire. In this context, having an efficient way to *a posteriori* identify actual parentage could be profitable both at the breeding scheme and the breeder levels. The objective of this study was to develop a SNP panel to accurately assign paternity in the main French goat breeds. About thirty animals have been genotyped with the GoatSNP50 chip in eight breeds: Alpine, Saanen, Angora, Corsican, Poitevine, Pyrénées, Fossés, Provençale. Four subsets of SNP (64 to 505) were theoretically defined based on call frequency, mendelian incompatibility, Hardy-Weinberg equilibrium test, Minor Allele Frequency and distribution along the genome. Then, the assignment rate of each SNP panel was estimated on Alpine and Saanen data from a QTL detection design, and by using a likelihood based approach. The 505 and 246 SNP panels showed the best results: the average MAF was about 0.42 for all the markers and populations, and the assignment rate was 100%. Before choosing the final panel, an assignment test will be performed in farm conditions (collection of blood samples, multiple-sired matings,…). A selection of SNP located in major genes of interest will be added to those for assignment in order to provide a useful tool to industry.

W529: International Goat Genome Consortium

An Update on Goat Genomics

Alessandra Stella, PTP Science Park, Lodi, Italy; ADAPTMAP Consortium, Lodi, Italy and Gwenola Tosser-Klopp, INRA, Castanet-Tolosan, France; International Goat Genome Consortium, CASTANET-TOLOSAN cedex, France

Goats are specialized in dairy, meat and fiber production, being adapted to a wide range of environmental conditions and having a large economic impact in developing countries. In the last years, there have been dramatic advances in the knowledge of the structure and diversity of the goat genome/transcriptome and in the development of genomic tools, rapidly narrowing the gap between goat and related species such as cattle and sheep. Major advances are: 1) publication of a *de novo* goat genome reference sequence; 2) Development of whole genome high density RH maps, and; 3) Design of a commercial 50K SNP array. Moreover, there are currently several projects aiming at improving current genomic tools and resources. An improved assembly of the goat genome using PacBio reads is being produced, and the design of new SNP arrays is being studied to accommodate the specific needs of this species in the context of very large scale genotyping projects (i.e. breed characterization at an international scale and genomic selection) and parentage analysis. As in other species, the focus has now turned to the identification of causative mutations underlying the phenotypic variation of traits. In addition, since 2014, the ADAPTmap project (www.goatadaptmap.org) has gathered data to explore the diversity of caprine populations at a worldwide scale by using a wide variety of approaches and data.

W530: International Goat Genome Consortium

Update on ADAPTMAP Working Groups and Discussion

Alessandra Stella, PTP Science Park, Lodi, Italy

W531: International Goat Genome Consortium

Conclusion

Gwenola Tosser-Klopp, INRA, Castanet-Tolosan, France

W532: International Phytomedomics and Nutriomics Consortium (ICPN)

Application of Phytomedomics and Nutriomics for Health Security

Chittaranjan Kole, Sam Higginbottom Institute of Agriculture, Technology & Sciences, Allahabad, India, Phullara Kole, C. Kole Foundation for Science and Society, Kalyani, India and Amit Dhingra, Department of Horticulture, Washington State University, Pullman, WA

W533: International Phytomedomics and Nutriomics Consortium (ICPN)

Deciphering the Genetic Control of Tomato Fruit Composition in Flavonoids in the Resequencing Era

Mathilde Causse[1], Laura Pascual[1], Christopher Sauvage[1], Elise Albert[1], Jean-Paul Bouchet[1], Dominique Brunel[2] and Marie-Christine Le Paslier[3], (1)INRA GAFL, Montfavet, France, (2)INRA - EPGV, Evry, France, (3)INRA, US1279 Etude du Polymorphisme des Génomes Végétaux, CEA-IG / Centre National de Génotypage, Evry, France

Identifying the genes controlling the variation of quantitative traits is a key goal for breeders. Genetic variations underlying quantitative traits (QTL) have been mapped by linkage mapping for years and positional cloning identified several QTLs. However linkage mapping is limited to the analysis of traits differing between two lines and the impact of genetic background on QTL effect has been underlined. Multi-allelic Advanced Generation Inter-Cross (MAGIC) populations allow a wide range of variability to be analyzed.

We have constructed a MAGIC population by crossing 8 tomato lines, representing a wide range of genetic diversity within the species. The large range of phenotypic variability represented by these lines was then characterized at different scales, metabolomic, proteomic and transcriptomic. The whole genomes of the 8 founder lines were resequenced identifying more than 4 millions SNPs when mapped onto the reference genome. A set of 1536 SNPs markers was then selected to genotype the MAGIC population and used to construct a linkage map. A large increase in recombination frequencies compared to bi-parental populations was shown. QTLs for fruit composition traits were mapped and related to the variations detected at the genome sequence and expression levels in the parental lines. QTL for flavonoid compounds were mapped and candidate genes identified will be presented.

W534: International Phytomedomics and Nutriomics Consortium (ICPN)

The Carrot Genome: A Framework to Study Health-Promoting Metabolite Accumulation

Massimo Iorizzo[1,2], Shelby Ellison[3], Douglas Senalik[4], Pim Satapoomin[2], Allen Van Deynze[5] and Philipp W. Simon[4], (1)Plants for Human Health Institute, Department of Horticulural Science, North Carolina State University,, Kannapolis, NC, (2)Department of Horticulture, University of Wisconsin-Madison, Madison, WI, (3)Department of Horticulture, University of Wisconsin, Madison, WI, (4)USDA-Agricultural Research Service, Vegetable Crops Unit, University of Wisconsin-Madison, Madison, WI, (5)University of California, Davis, CA

Carrot (*Daucus carota* subsp. *carota* L.) is among the top ten global vegetable crops and root crops both in terms of production and market value. Carrot is well recognized for its health properties which can be attributed to primary and secondary metabolites such as carotenoids, terpenoids, and anthocyanins. Metabolite accumulation represents an important trait in carrot production since it determines its nutritional value, color, taste and plays a critical role in human health. Despite their importance in carrot production little is known about the genetic mechanism controlling primary and secondary metabolite accumulation in carrot roots. Our group has led the development of the first chromosome scale assembly and characterization of the carrot genome. The assembly covers ~90% (422.3 Mb) of the estimated genome size (473 Mb) with an N50 of 12.7 Mb. Sixty superscaffolds covering 85.6% of the assembled genome were anchored to the nine pseudomolecules, containing over 95% of predicted genes. Genome characterization included a curated annotation of genes involved in the isoprenoid and flavonoid pathways and over 3,000 genes involved in regulatory function, including those controlling anthocyanin accumulation. A consensus linkage map integrating over 2,000 markers

and 15 qualitative and quantitative traits, mainly related to metabolite accumulation, has been developed and anchored to the physical map. Integration of QTL mapping, transcriptome and genomic information are being used to identify candidate genes for metabolite accumulation.

W535: International Phytomedomics and Nutriomics Consortium (ICPN)

A Candidate Gene Association Approach Combined with QTL Mapping Reveals the Genetic Architecture of Glucosinolate Content in *Brassica rapa* Leaves

Yong Pyo Lim, Xiaonan Li, Su-Ryun Choi, Vignesh Dhandapani and Wenxing Pang, Chungnam National University, Daejeon, South Korea

Glucosinolates are sulfur-rich secondary metabolites mainly found in *Brassicacea* family. GSLs and their degradation products plays an important role in human health, pathogen and insect prevention, as well as in special flavors and tastes. We firstly performed a conventional QTL analysis using a *B. rapa* segregated population combined with candidate gene association approach by using natural population in order to identify the genomic region and genes regulating glucosinolates biosynthesis in *B. rapa* crops. Both QTL and association mapping result revealed that paralogous of R2R3 MYB transcription factor, MAM gene family and BCAT-4 on A02, A03, A04, A07 syntenic region regulated the individual, three types and total aliphatic glucosinolate with gene redundancy function, guaranteeing the effective biosynthesis and metobolisam of glucosinolate. All of these associated genes had a high level of linkage disequilibrium in *B. rapa* genome. Moreover, two haplotypes forming by four SNPs in *B. rapa* MYB28b gene were significant related to glucosinolate content. Haplotype II with minor allele frequency in *B. rapa* could be used for selecting high gluconapin, aliphatic and total glucosinolate content.

W536: International Phytomedomics and Nutriomics Consortium (ICPN)

Understanding Genetic Controls of Capsaicin Content in *Capsicum annuum* L

Umesh K. Reddy and Padma Nimmakayala, Department of Biology, West Virginia State University, Institute, WV

Accumulated capsaicinoid content and increased fruit size are traits resulting from *Capsicum annuum* domestication. In this study, we used a diverse collection of domesticated and wild *C. annuum* to generate 66,960 SNPs using genotyping by sequencing. Principal component analysis and identity by state were used in a mixed linear model of capsaicin and dihydrocapsaicin content and fruit weight to reduce spurious associations because of confounding effects of subpopulations in genome-wide association study (GWAS). Selfed accessions were grown in three replications during two seasons (2011 and 2012). GWAS revealed 14 SNPs commonly associated with capsaicin and dihydrocapsaicin content and 15 associated with fruit weight in both years. Five associated SNPs for capsaicin and three for fruit weight were nonsynonymous and significant after correction for false discovery rate and were homologous to known fruit-development genes in tomato and other plants. When scanning pairwise fixation index distribution of domesticated and wild accessions across the genome, we identified a segment of 177 Mb on chromosome 11 that was under strong selection sweep. Of 659 genes located in this sweep area, 30 were under high linkage disequilibrium, with reduced nucleotide diversity levels. Annotation of the genes in the sweep coupled with GWAS revealed their important roles in domestication.

W537: International Phytomedomics and Nutriomics Consortium (ICPN)

Characterization of unknown plastid targeted genes involved in production of metabolic compounds in medicinally active plants

Amit Dhingra, Department of Horticulture, Washington State University, Pullman, WA and Ryan Christian, Molecular Plant Sciences Graduate Program, Washington State University, Pullman, WA

Plants have been used as medicine throughout human history. However, many of the medicinally relevant compounds are synthesized via pathways that are only partially characterized. In order to maximize the efficient production of these valuable compounds in their native species or in a heterologous plant species, elucidation of the genes involved in these pathways is key. For many of these pathways, key steps are known or suspected to occur within the highly reductive environment of the plastid. Precursors such as terpenes and aromatic amino acids are also produced in plastids, giving this organelle an unparalleled degree of control over biochemical flux. Production of the antimalarial drug quinine (*Cinchona pubescens*), the neurotransmitter antagonist atropine (*Datura innoxia*), and the anticancer drug paclitaxel (*Taxus x media*) are known to be largely plastid-synthesized. We are characterizing the predicted plastid-targeted proteomes of these species using clustering algorithms to elucidate common and unique genes from each species. We analyzed RNAseq data from these species using a custom assembly and analysis workflow. The resulting assemblies were used to derive peptide sequences that were further analyzed using TargetP to identify plastid targeted genes. Comparative analysis of predicted plastid targeted proteomes was performed using a recently published approach (Schaeffer et al., 2014). This analysis is expected to yield gene candidates involved in respective medicinal compound biosynthesis, and will facilitate the unraveling of the biochemical pathways involved in the synthesis of these drugs.

W538: International Rice Informatics Consortium

Update on the Rice Informatics Consortium: How to work together

Nickolai Alexandrov, International Rice Research Institute, Los Banos, Laguna, Philippines

Five goals were set by the Advisory Committee for the development of the International Rice Informatics Consortium for 2015: 1) develop a mechanism for unified germplasm tracking, 2) select subset(s) of germplasm for detailed studies and organize corresponding metadata using existing ontologies, 3) create a curated database of functional loci annotation, 4) establish communication with similar plant consortia (Arabidopsis, wheat, soy, etc), and 5) approach funding agencies. We will discuss our achievements and priorities for the coming year. Additional members have joined, and we have established new collaborations. Several collaborative projects are being developed within IRIC where we are testing software development frameworks for making such collaborations efficient and sustainable.

W539: International Rice Informatics Consortium

Exploring the Rice Dispensable Genome using a Metagenome-like Assembly Strategy

Weibo Xie, Huazhong Agricultural University, Wuhan, China

The dispensable genome of a species, consisting of the dispensable sequences present only in a subset of individuals, is believed to play important roles in phenotypic variation and genome evolution. However, construction of the dispensable genome is costly and labor-intensive at present, and so the influence of the dispensable genome in genetic and functional genomic studies has not been fully explored. We construct the dispensable genome of rice through a metagenome-like de novo assembly strategy based on low-coverage (1-3×) sequencing data of 1483 cultivated rice (*Oryza sativa*L.) accessions. Thousands of protein-coding genes are successfully assembled, including most of the known agronomically important genes absent from the Nipponbare rice reference genome. We develop an integration approach based on alignment and linkage disequilibrium, which is able to assign genomic positions relative to the reference genome for more than 78.2 % of the dispensable sequences. We carry out association mapping studies for rice grain width and 840 metabolic traits using 0.46 million polymorphisms between the dispensable sequences of different rice accessions. About 23.5 % of metabolic traits have more significant association signals with polymorphisms from dispensable sequences than with SNPs from the reference genome, and 41.6 % of trait-associated SNPs have concordant genomic locations with associated dispensable sequences. Our results suggest the feasibility of building a species' dispensable genome using low-coverage population sequencing data. The constructed sequences will be helpful for understanding the rice dispensable genome and are complementary to the reference genome for identifying candidate genes associated with phenotypic variation.
Reference: Yao W, Li G, Zhao H, Wang G, Lian X, Xie W*. Exploring the rice dispensable genome using a metagenome-like assembly strategy. *Genome Biology*, 2015, 16: 187

W540: International Rice Informatics Consortium

Bioinformatic Approaches for Comprehensive Variants Discovery on the 3,000 Rice Genomes Project

Jorge A. Duitama Castellanos, International Center for Tropical Agriculture (CIAT), Cali, Colombia

The 3,000 rice genomes project (3KRGP) is arguably the largest current sequencing effort to elucidate genomic variability in a staple crop. Information obtained from bioinformatic analysis of whole genome resequencing (WGS) of the 3,000 varieties selected for this project will make a great resource not only for rice biology and population dynamics but also for applied rice breeding. To maximize the information obtained from WGS data we recently developed the software package NGSEP, which combines state-of-the-art algorithms to identify and genotype different kinds of genomic variation including SNPs, SSRs, indels and copy-number variants (CNVs). We demonstrated the use of NGSEP to reveal millions of SNPs and thousands of structural variation events on a panel of 94 *O. sativa* varieties an 10 wild relatives, which included 54 elite cultivars both from North America and Latin America. As part of the IRIC consortium, we now have used the algorithms implemented in NGSEP to predict structural variants in the complete WGS dataset of the 3KRGP. We identified more than one thousand CNVs and 10 thousand large deletions with alternative allele frequencies greater than 0.01 in regions with low repeat density and covering 4.5 and 1 Mbp of the genome respectively. About 50% of these events span, and sometimes even completely cover annotated genes, suggesting significant functional variation within rice due to structural variants. We expect that the variants obtained by this analysis complement the SNPs previously released to assemble a comprehensive information resource useful for several groups across the rice research community.

W541: International Rice Informatics Consortium

Sequencing and Assembly of the Rice Variety N22 (Aus Group): a New Reference Genome to Study Comparative, Evolutionary and Functional Genomics of Rice

David Kudrna[1], Dario Copetti[1], Maria Elizabeth B. Naredo[2], Sheila Mae Q. Mercado[3], Kenneth L. McNally[3] and Rod A. Wing[1], (1)Arizona Genomics Institute, University of Arizona, Tucson, AZ, (2)International Rice Research Institute, Metro Manila, Philippines, (3)T.T. Chang Genetic Resources Center, International Rice Research Institute, Los Banos, Philippines

Since the release of 3000 resequenced rice genomes in May 2014, the global effort to drill deeper into this valuable data has begun to materialize. Current analyses can group the 3K set into 9-15 subpopulations. The goal of the international community is to genetically and phenotypically characterize natural variation that exists within IRRI's 125K GeneBank accessions to address the most relevant traits for rice production, such as biotic and abiotic stress tolerances, and yield and grain quality. In support of this goal we plan to fully sequence and annotate the genomes and transcriptomes of 3-4 accessions from each subpopulation and map the 3K data to each genome for variation discovery and GWAS and genomic selection studies.
Here we present the first high-quality genome sequence of an aus variety rice - Nagina22. N22 is a reference upland adapted rice. We produced and assembled 65 genome equivalents of PacBio long-read data that resulted in an assembly of 373Mb (1519 contigs, contig N50 of 906Kb, longest contig 3.5Mb). Contigs were assigned to pseudomolecules after alignment to the *O. sativa ssp. japonica* RefSeq thus providing a high quality reference sequence for aus rice. Genome annotation and comparative analyses define features of this variety and importantly provide a reference for the alignment of the resequencing data from the rice 3K project. With additional high-quality reference genome assemblies, the goal of characterizing and classifying the unlocked genetic potential of this vital crop will facilitate the next generation of green super crops to help feed the world.

W542: International Rice Informatics Consortium

Sharing Experience: What Can We Learn from Each Other Developing Plant Informatics Systems

David Edwards, University of Western Australia, Perth, Australia

Several groups are each developing plant specific information systems to help with the management and interrogation of the growing abundance and diversity of available data. Some aspects of these systems will be species specific though many will be applicable across species.
The aim of this discussion is to explore ways for groups to share their experience, avoid duplicated effort and work towards standards for plant data, working across species and expert groups.

W543: International Rice Informatics Consortium

Open Discussion for Community Input into IRIC Development

Kenneth L. McNally, T.T. Chang Genetic Resources Center, International Rice Research Institute, Los Banos, Philippines

In this session, your input is wanted into how we further develop our community, what tools we create and/or deploy and what data resources should be prioritized for curation. This builds on prior discussion from previous meetings summarized in the document at:
http://iric.irri.org/resources/downloadable-files/ISRFG12_community_input.docx
Please take a look at this document and come prepared for discussion.

W544: International Sheep Genomics Consortium
Reference Genome Sequence Updates: Texel Improvements and Rambouillet Progress
Kim C. Worley, Baylor College of Medicine, Houston, TX

W545: International Sheep Genomics Consortium
Building the LD Chip and an Update on the Sheep Genomes Database
Rudiger Brauning[1], Antonello Carta[2], Carlos Gabriel Ciappesoni[3], Shannon Clarke[1], Noelle Cockett[4], Christine Couldrey[5], Hans D. Daetwyler[6], Michael P. Heaton[7], James Kijas[8], Denis Larkin[9], Alan McCulloch[1], John McEwan[1], Sean McWilliam[10], Carole R. Moreno[11], Suzanne Rowe[1], Gary Saunders[12] and Ricardo Ventura[13], (1)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)Istituto Zootecnico e Caseario per la Sardegna, Sassari, Italy, (3)National Institute of Agriculture Research, Rincon del Colorado, Canelones, Uruguay, (4)Utah State University, Logan, UT, (5)LIC, Hamilton, New Zealand, (6)Department of Economic Development, Jobs, Transport & Resources, Bundoora, Australia, (7)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE, (8)CSIRO Agriculture Flagship, St. Lucia, Australia, (9)Aberystwyth University, London, United Kingdom, (10)CSIRO Agriculture QBP, St. Lucia, Australia, (11)INRA, Toulouse, France, (12)EMBL-EBI, Cambridge, United Kingdom, (13)Beef Improvement Opportunities, Guelph, ON, Canada
The LD chip is a new addition to the family of ovine SNP ships. Here we present its final design. The major driver was a low cost chip for genomic selection in industry. We designed and tested it for imputation. In addition to imputation SNPs the chip also has parentage and key functional SNPs and has backwards compatibility with earlier versions of ovine LD chips. It is the first LD chip to fully utilise version 3.1 of the genome as well as results derived from the HD ovine chip.
The SheepGenomesDB is an electronic warehouse containing sequence variants called from the expanding collection of sheep genomes being generated across our research community. Through the application of a single harmonised pipeline for read QC, mapping, variant detection and annotation, SheepGenomesDB makes available variant collections derived in a standardised manner supported by the European Variation Archive. The first analysis run has been completed, with read mapping and variant calling performed for more than 250 sheep genomes from over 50 breeds. This identified in excess of 25 million SNP. The creators view SheepGenomesDB as an essential resource for researchers interested in gene mutation discovery, imputation as it relates to genome-wide association and genomic prediction as well as investigations into genome evolution, domestication and the consequences of selection.

W546: International Sheep Genomics Consortium
RNA Sequencing of 150 Lambs Connects Meat Phenotypes with Gene Expression
Hans D. Daetwyler[1,2], Bolormaa Sunduimijid[1], Ralph Behrendt[3], Matthew Knight[3], Lysandra Slocombe[3], Brett Mason[1], Claire Prowse-Wilkins[1], Christy J Vander Jagt[1], Ben J Hayes[1,2] and Amanda J. Chamberlain[1], (1)DEDJTR, Bundoora, Australia, (2)La Trobe University, Bundoora, Australia, (3)Agriculture Research, DEDJTR, Hamilton, Australia
Gene expression analysis can aid in prioritising regions or classes of variants for genomic prediction and help increase the understanding of quantitative traits. Liver and muscle tissue were collected at slaughter for 150 lambs, with up to 70 meat traits recorded per lamb. Their dams were managed to high, medium, and low body condition scores (BCS) during mid-to-late pregnancy with the lambs finished on three different diets. Differential expression of genes (DEG) investigated contrasting tissue, BCS, lamb diets, other treatment differences, as well as high and low lamb carcass eye muscle width (CEMW). Covariance matrices of depth normalized read counts per exon were fitted in a mixed model to estimate the proportion of the phenotypic variance explained by co-expression. Expression QTL (eQTL) analyses were performed on the Ovine high density SNP chip and on the newly discovered SNP using RNA sequence. A large number of DEG were identified between tissues, but only the low versus high BCS comparison resulted in any DEG for treatments. DEG was also found for lambs with high and low CEMW. A strong trend toward down regulation was observed in all tests, except for BCS, where all DEG were overexpressed in lambs from ewes with higher condition score during mid-late pregnancy. Co-expression accounted for up to 79% of the phenotypic variance but was very variable across traits. Many eQTL passed multiple testing and are now being compared to results from an independent meat trait genome-wide association study of 10,000 lambs.

W547: International Sheep Genomics Consortium
Towards a Methylation Chip and the Way Forward for Genotyping
Shannon Clarke[1], Rudiger Brauning[1], Ken G Dodds[1], Christine Couldrey[2], Tracey Van Stijn[1], Rayna Anderson[1] and John McEwan[1], (1)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand, (2)LIC, Hamilton, New Zealand
Recent years have seen impressive progress in the ability to improve the rate of genetic gain through the use of SNP chips. However, the expression of a gene is not solely defined by the DNA sequence, and it is becoming clear that simply knowing nucleotide changes does not provide the full picture for predicting downstream animal production. Over the last decade, the importance of epigenetics (including DNA methylation) in regulating gene expression and phenotype has become apparent, even though mechanisms of action remain elusive. Here we report our progress on the development of an Illumina Infinium Methylation chip for sheep.
In addition, to developing a suite of array based genotyping tools, AgResearch has also invested in genotyping by sequencing (GBS) methods. High-throughput GBS methodology produces SNP genotypes that are supported by varying depth of sequence reads, dependent on the number of samples and proportion of the genome assayed within a lane of sequencing. A recently-developed algorithm produces bias free genomic relationship matrices, based on allele read depths, which can be interrogated to estimate: breed composition, pedigree, traceability, inbreeding

and co-ancestry as well as be included directly in existing mixed models to estimate breeding values. The GBS pipeline is cost competitive compared to array based technologies and offers the potential to eliminate ascertainment bias. Its disadvantage is the requirement for high quality DNA, however, this can be overcome by appropriate sample collection, storage and extraction methods at minimal cost.

W548: International Sheep Genomics Consortium

The transcriptomic and regulatory dynamics of the rumen epithelium of sheep

Ruidong Xiang[1], Jody McNally[2], Suzanne Rowe[3], Arjan Jonker[3], Cesar Pinares-Patino[3], Alan L. Archibald[4], Jude Bond[5], V. Hutton Oddy[5], Phil Vercoe[6], John McEwan[3] and Brian Dalrymple[7], (1)CSIRO Agriculture, St Lucia, Australia, (2)CSIRO Agriculture, Armidale, Australia, (3)AgResearch Ltd, Invermay Agricultural Centre, Mosgiel, New Zealand, (4)The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom, (5)DPI NSW, Armidale, Australia, (6)The University of Western Australia, Crawley, Australia, (7)CSIRO Agriculture, St. Lucia, QLD, Australia

The rumen is the interface between the diet, rumen microbes and the animal. However a fundamental mechanistic understanding of the control of the rumen epithelial interactions is lacking. We performed comparative transcriptomic analysis of the whole rumen wall of sheep measured for methane production and other rumen-based traits in two different datasets, including one with four combinations of amount and composition of diet. Few genes from the muscle layer were significantly responsive to diet. In contrast genes involved in cell cycle (proliferation of epithelial cells) and rumen metabolism (also in the epithelial layer) were the most dynamic signals and responded to different diets. The majority of gene expression and phenotypic variation was explained by feed consumption level. Using a network approach metabolic genes were further separable based on general cellular (electron transport and intracellular transport) and rumen-specific (keto-acids and lipids) processes. Enrichment of transcription factor binding sites in genes with correlated expression identified potential key transcription factors actively expressed in the rumen epithelial layer with roles in regulating epidermal cell division and differentiation. Many of the transcription factors identified have characterised roles in skin growth and similar expression patterns in human skin and sheep rumen relative to other genes. It appears likely that the control of rumen growth is a combination of generic epithelial processes (such as in the skin) and rumen specific processes. Our results demonstrate the promise of transcriptomics to elucidate the mechanisms of regulation of rumen epithelial growth and interaction with diet and the microbial populations.

W549: International Wheat Genome Sequencing Consortium (IWGSC)

Assembly and Validation of the Wild Emmer Wheat Genome

Assaf Distelfeld, Faculty of Life Sciences - Tel Aviv University, Tel Aviv, Israel

Wild emmer wheat (*Triticum turgidum* ssp. *dicoccoides*, genome BBAA) gene-pool is an important source for wheat research and improvement. The International **W**ild **E**mmer **W**heat Genome **Seq**uencing consortium (**WEWseq** - http://wewseq.wix.com/consortium) aims to explore this resource by providing cost-effective high quality genomic sequence. As a step towards achieving genome assembly of tetraploid wheat we hybridized durum wheat (*T. turgidum* ssp. *durum*, cv. Svevo) with wild emmer wheat and developed a recombinant inbred line (RIL) population. An ultra-dense genetic map was constructed by combing data from the wheat 90K iSelect SNP genotyping assay along with genotyping by sequencing (GBS) data. Next, we have performed paired-end (2 x 260 nucleotides) and mate-pair shotgun sequencing equal to about 190x genome coverage of the wild emmer accession 'Zavitan'. The short reads were assembled to scaffolds using NRGene's DeNovoMAGIC^{TM}2.0 assembler and cover ~90% of the genome with L_{50} of 7 Mb and N_{50} of 414. Most of the scaffolds were anchored to our genetic map more than 71,000 gene models were annotated. The integrity and quality of the genome assembly were assessed using different approaches and the latest results will be presented.

W550: International Wheat Genome Sequencing Consortium (IWGSC)

IWGSC whole genome shotgun sequencing of Chinese Spring: Towards a Reference Sequence of Wheat

The International Wheat Genome Sequencing Consortium, IWGSC, Lee's Summit, MO, **Curtis J Pozniak**, University of Saskatchewan, Saskatoon, SK, Canada, Andrew G. Sharpe, National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada, Jesse Poland, Kansas State University, Manhattan, KS, Mike Thompson, Illumina, Inc, River Falls, WI, Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany, Assaf Distelfeld, Tel Aviv University, Tel-Aviv, Israel, Gil Ronen, NRGene Ltd., Ness Ziona, Israel, Frédéric Choulet, INRA GDEC, Clermont-Ferrand, France, Kellye Eversole, Eversole Associates, Bethesda, MD and Jane Rogers, International Wheat Genome Sequencing Consortium, Cambridge, United Kingdom

Wheat is one of the world's most important food crops and provides >20% of the protein and calories for the world's population, but it is the only major crop that lacks an ordered genome sequence. The lack of a reference sequence has slowed identification of genes underlying phenotypic expression of agriculturally important traits. The IWGSC has followed an approach to develop separate physical maps, draft sequences, and high quality sequencing of the Minimum Tiling Paths (MTP) of mapped BAC clones for each chromosome to accelerate the completion of the gold standard reference sequence. Recent innovations in Illumina sequencing technology and NRGene's computational genomics have enabled the IWGSC to produce rapidly a whole genome shotgun sequence of the "Chinese Spring" hexaploid wheat genome to complement ongoing BAC-based sequencing strategies. We assembled Illumina generated short reads using NRGene's DeNovoMAGICTM2.0 assembler to span 14.6 Gb of the 17 Gb genome. We are currently validating and anchoring the assembly using IWGSC resource. We will present our results and discuss utilization of this new resource as an additional tool to complete the reference sequence for all 21 wheat chromosomes.

W551: International Wheat Genome Sequencing Consortium (IWGSC)

BioNano Genome Map of Bread Wheat 7DS Arm Supports Sequence Assembly and Analysis

Hana Simkova[1], Helena Stankova[1], Alex Hastie[2], Saki Chan[2], Jan Vrana[1], Zuzana Tulpova[1], Paul Visendi[3], Satomi Hayashi[3], Ming-Cheng Luo[4], Jacqueline Batley[5], David Edwards[6] and Jaroslav Dolezel[1], (1)Institute of Experimental Botany, Olomouc,

Czech Republic, (2)BioNano Genomics, San Diego, CA, (3)University of Queensland, Brisbane, Australia, (4)Department of Plant Sciences, University of California, Davis, Davis, CA, (5)University of Western Australia, Crawley, Australia, (6)University of Western Australia, Perth, Australia
Optical mapping and its modification - genome mapping in nanochannel arrays (BioNano mapping) - provides a complementary tool to next-generation sequencing technologies. It generates maps of a short sequence motif along DNA stretches of hundreds to thousands kb in length. The BioNano technology excels in accurate length measurement and high throughput, which makes mapping gigabase-sized genomes feasible. Still, the 17-Gb hexaploid genome of bread wheat appears too complex to be tackled as a whole. Moreover, as the wheat reference genome sequence is being produced by sequencing physical maps of individual chromosomes/arms, it seems practical to follow the chromosome-based strategy of IWGSC and produce BioNano maps from individual chromosomes. Here we chose the short arm of chromosome 7D (7DS) as a model to demonstrate that it is possible to couple chromosome flow sorting with BioNano mapping to create a de novo genome map.
We have constructed a high-resolution map of chromosome 7DS composed of 371 contigs with an N50 of 1.3 Mb. Long DNA molecules obtained from flow-sorted chromosomes facilitated chromosome-scale analysis of repetitive sequences and revealed a ~800-kb array of tandem repeats intractable to current DNA sequencing technologies. Anchoring 7DS sequence assemblies obtained by clone-by-clone sequencing to the 7DS BioNano map provided a valuable tool to improve the BAC-contig physical map and validate sequence assembly. Moreover, a BioNano map was also constructed from 7DS arm flow-sorted from wheat line CI2401, which enabled characterizing structural variation between two wheat accessions at chromosome-arm level.

W552: International Wheat Genome Sequencing Consortium (IWGSC)

Genome Sequence of *T. uartu*

Hong-Qing Ling, Institute of genetics and developmental biology, CAS, Beijing, China

Genome sequencing of *Triticum urartu* – the progenitor of wheat A genome

Hong-Qing Ling, Hui Liu, Biao Ma[1], Hua Sun[1], Shancen Zhao[2], [3]Michiel van Eijk, Zhensheng Li[1], Aimin Zhang[1], Daowen Wang[1], Chengzhi Liang[1]
[1]*Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China.* [2]*BGI Shenzhen, Shenzhen, China*. [3]Keygene, Wageningen, The Netherlands, *E-mail: hqling@genetics.ac.cn*
Bread wheat (*Triticum aestivum*), one of the most widely cultivated and consumed food crops in the world, is a hexaploid containing A, B and D genomes. Due to its complex polyploidy nature and large genome size (17 Gb), the genetic and functional analysis of bread wheat is extremely challenging. *Triticum urartu* is the progenitor of wheat A genome. It plays a central role in wheat evolution, domestication and genetic improvement because the A genome is the basic genome of bread wheat and other polyploidy wheats. Therefore, we are working on the genome sequencing of *T. urartu* since 2009.
Two years ago, we have sequenced *T. urartu* accession G1812 using a whole-genome shotgun strategy on the Illumina HiSequation (2000) platform, and generated a draft genome of wheat A genome (Ling et al., 2013, Nature 496: 87-90). To completely sequence the wheat A genome, we constructed 3 BAC libraries and created BAC contigs using the whole genome profiling technology by collaboration with Keygene, the Netherlands. Based on BAC contig data, we selected 47,223 BACs by MTP approach, and sequenced them. Additionally, we also generated more than 80 G sequence data by whole genome sequencing using the third generation sequencing technology (PacBio RS II). The sequence assembly has been finished and the genome analysis is in progress. For assignment and order of the genome sequences on their corresponding chromosome, we sequenced 480 F2 plants derived from two *T. urartu* accessions (G1812 and G3146) using RAD-sequencing technology, and identified more than 400,000 SNPs. With them, a high resolution SNP map has been constructed. In conclusion, completing the genome sequence of *T. urartu* will provide a diploid reference for the analysis of polyploidy wheat genomes, and is a valuable resource for the genetic improvement of wheat to meet the future challenges of global food security and sustainable agriculture.

W553: International Wheat Genome Sequencing Consortium (IWGSC)

Assembly of the 4.5Gb Ancestral Wheat D-Genome from Hybrid PacBio and Illumina Data

Aleksey Zimin, University of Maryland, College Park, MD
The developments in DNA sequencing technology over the past several years have enabled large number of scientists to obtain sequences for the genomes of their interest at a fairly low cost. Illumina Sequencing was the dominant whole genome sequencing technology over the past few years due to its low cost. The Illumina reads are short (up to 300bp) and thus most of those draft genomes produced from Illumina data are very fragmented which limits their usability in practical scenarios. Longer reads are needed for more contiguous genomes. Recently Pacbio sequencing made significant advances in developing cost-effective long-read (>10000bp) sequencing technology and their data, although several times more expensive than Illumina, can be used to produce high quality genomes. Pacbio data can be used for de novo assembly, however due to its high error rate high coverage of the genome is required this raising the cost barrier. A solution for cost-effective genomes is to combine Pacbio and Illumina data leveraging the low error rates of the short Illumina reads and the length of the Pacbio reads.We implemented this solution in MaSuRCA assembler 3.2.1, available as alpha-release now. We demonstrate superiority of our algorithm to the other published hybrid techniques and show the results of applying our assembler to the assembly of the ~4.5Gbp genome of the ancestral wheat, A. tauschii.

W554: International Wheat Genome Sequencing Consortium (IWGSC)

Exploitation of the 5BS Physical Map to Complete the SKr Crossability Locus

Véronique Lesage[1], Frederic Choulet[2], Sonia Vautrin[3], Elena A. Salina[4], Marie-Claire Debote[1], Bouzid Charef[1], Hélène Bergès[3], Catherine Feuillet[5] and **Pierre Sourdille**[2], (1)INRA - UBP UMR 1095 GDEC, Clermont-Ferrand, France, (2)INRA GDEC, Clermont-Ferrand, France, (3)INRA - CNRGV, Castanet Tolosan, France, (4)Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia, (5)Bayer CropScience, Morrisville, NC
Most elite wheat varieties cannot be crossed with related species thereby restricting greatly the germplasm that can be used for alien introgression in breeding programs. Inhibition to crossability is controlled genetically and a number of QTL have been identified to date, including *SKr*, a strong QTL affecting crossability between wheat and rye located at the distal end of chromosome 5BS. We started the positional cloning of this

gene and we elaborated a first draft of a physical map in both the crossable and non-crossable lines. We used the newly developed physical map of the 5BS chromosome arm to partly fill in the gaps and complete the sequence. Sequence was made of three main contigs and covered ~1 Mb. The sequence was used to develop several tens of markers useful for the screening of a mutant population derived from irradiated seeds of the cultivar Renan. This led to the identification of seven mutants showing crossability while Renan is a non-crossable variety. This allowed the refining of the positional cloning of the gene and reduced the number of candidate genes. Validation of these genes will be done through transgenesis (overexpression and knock-out) and through EMS mutant screening.

W555: International Wheat Genome Sequencing Consortium (IWGSC)

Early Career Speaker: Evolution of Genome Structure in Polyploid Wheat Revealed by Comparison of Wheat and *Aegilops tauschii* Whole-Genome BioNano Maps

Tingting Zhu[1], Juan C. Rodriguez[1], Karin R. Deal[2], Sonny Van[1], Jan Dvorak[2] and Ming-Cheng Luo[1], (1)Department of Plant Sciences, University of California, Davis, CA, (2)Department of Plant Sciences, University of California, Davis, Davis, CA

It has been suggested that changes in genome structure and expression played an important role in the evolution of hexaploid wheat (genomes AABBDD). Hexaploid wheat originated about 8,000 years ago via hybridization of tetraploid wheat (genomes AABB) with diploid *Aegilops tauschii* (genomes DD). Because of this recent origin, it may be possible to detect and analyze in detail structural changes that have taken place in the wheat D genome. Genome-wide BioNano genome (BNG) maps are an exceptionally powerful tool for genome comparisons. We discuss the construction of BNG maps for *Ae. tauschii* and wheat and their subsequent comparison for the detection of structural differences between the D genome of wheat and that of *Ae. tauschii.* These comparisons revealed that a vast majority of these changes are indels. The BNG map comparisons can identify those that have taken place during wheat evolution, precisely locate them, and assess the role of recombination and location on the centromere-telomere axis in their origin. This work is a part of the NSF-funded Project IOS-1238231 to generate a reference sequence for the genome of *Ae. tauschii* (http://aegilops.wheat.ucdavis.edu/ATGSP/).

W556: iPlant Education: Genomics, DNA Barcoding, RNA-Seq, and Data Science for the Undergraduate Classroom

Using DNA Barcoding and Educational Bioinformatics to Create Authentic Research Experiences for Science Students

Stephen Harris, Columbia University, New York, NY

In a pursuit to teach enough content to our science students, we often forget to instill the skills more generally characteristic of the scientific approach to inquiry. Here, we developed an experiential science curriculum that introduces students to modern biological research. Traditional means of studying biodiversity depend on expert knowledge from individuals with years of education and training. Recent techniques like DNA barcoding, the process of identifying species based on short fragments of DNA, can be used to quickly identify species and to provide easy access to taxonomic information. We use DNA barcoding as the foundation of a research course that serves as an alternative to more traditional laboratories, which often have known outcomes and lack student-generated investigations. In order to gain a deeper understanding of the habits of a scientific mind, students pursue individual research projects and their data is managed and analyzed using the DNA Subway bioinformatics platform. Students work in small groups, develop a research question or identify a species of interest, collect samples from their local environments, and then perform all related laboratory work and analysis. Depending on the laboratory facilities available to classrooms, students can perform DNA extraction, gene amplification using PCR, and gel electrophoresis to confirm results. Samples are then mailed to a third party facility for sequencing. Once sequences are in hand, students can use the DNA subway pipeline developed by the iPlant Collaborative at Cold Spring Harbor Laboratory to edit, align, and identify results. At the final stages of analysis, DNA subway provides the most important component of our DNA barcoding curriculum, the ability to upload results directly to GenBank, the National Institute of Health's genetic sequence database, to be made available to the professional scientific community. We have successfully implemented this curriculum in a variety of settings, including high school classrooms in NYC and workshops for undergraduate students in Belize. By introducing students to cutting-edge research techniques, our inquiry-based curriculum promotes evolution education, increases interest in conservation and biodiversity, and allows students to contribute directly to the scientific community.

W557: iPlant Education: Genomics, DNA Barcoding, RNA-Seq, and Data Science for the Undergraduate Classroom

RNA-Seq in the Classroom: Pathways to Undergraduate Research

Carrie Thurber, Abraham Baldwin Agricultural College, Tifton, GA

Modern biology students will be expected to use and analyze whole-genome and whole-transcriptome data sets across the medical, agricultural and environmental fields. I introduce sophomore and junior level undergraduates to these kinds of data in my introductory genetics lecture and lab course. Students spend one to two, 3-hour lab periods working through an online tutorial using the DNA Subway Green Line and analyze publically available RNA-Seq data. Not only can I assess their understanding and ability to think critically about these kinds of data, I also generate interest in further independent student research with the goal of publication and presentation. My current research student is working on a project involving *C. elegans* neural development.

W558: iPlant Education: Genomics, DNA Barcoding, RNA-Seq, and Data Science for the Undergraduate Classroom

Sifting Through Metagenomes using the iPlant Discovery Environment

Carlos C. Goller, North Carolina State University, Raleigh, NC

Teaching undergraduate and graduate students about metagenomics and analyzing sequence data can be challenging when participants are not familiar with the software. The iPlant Discovery Environment provides a user-friendly yet powerful platform to begin to perform metagenomic analyses in a classroom setting. Since the fall of 2013, the 8-week metagenomics course at the North Carolina State University Biotechnology Program has used the iPlant Discovery Environment to introduce students to cloud-based computing. A series of computer activities based on worksheets familiarize students with procedures to download sequences from databases, assess the quality of next-generation sequencing reads, launch applications to analyze sequences to infer taxonomy, and assemble metagenomes. Students learn key procedures in an environment that is user-friendly. Sharing of large sequence files and results with groups of students allows instructors to easily manage laboratory sessions. Pairs of students are then tasked with diagramming the procedures they followed to reinforce the steps for each workflow and propose alternatives.

Combined with reflective questioning, diagramming helps students break down complex processes into simple steps. The iPlant Discovery Environment offers an increasing number of tools for analyses of next-generation sequencing data, genomics, and metagenomics on a platform that is user-friendly, powerful, and promotes interaction among student groups.

W559: iPlant Education: Genomics, DNA Barcoding, RNA-Seq, and Data Science for the Undergraduate Classroom

Data Carpentry: Data Skills Training to Enable More Effective Research

Tracy Teal, Data Carpentry, Davis, CA

Our increasing capacity to collect data is changing science. This is particularly true in genomics where, with high-throughput sequencing, data production is no longer a bottleneck. There is great potential for discovery, but we are primarily failing to translate this sea of data into scientific advances, because researchers are not trained in the skills needed for effective management and analysis. The question then becomes, in addition to scaling data production and computation, how do we develop and deliver training to scale data literate researchers? Course curriculums are slow to change, need qualified instructors and are already full. Short courses are oversubscribed and reach a limited number of participants. To provide scalable and distributed training, Data Carpentry develops and teaches domain-specific hands-on workshops in data organization, management, and analysis. This is a grassroots training effort developed by practitioners for practitioners, who identify core skills and collaboratively develop lessons. All lessons are open source, and workshops are taught by volunteers trained by the Software Carpentry Foundation. A particular focus has been on lessons for working with genomics data, designed for people with little to no prior computational experience. The workshop teaches in two days how to organize genomic data and metadata, use cloud computing, use the command line to run bioinformatics pipelines and analyze and visualize data in R – the full data lifecycle. Workshops are in high demand, but this model allows for scaling of training and teaches the foundational skills to get biologists started managing and analyzing their data effectively.

W560: iPlant Education: Genomics, DNA Barcoding, RNA-Seq, and Data Science for the Undergraduate Classroom

Creating a Positive Space to Train Native Americans / American Indians in Genomics

Joslynn Lee, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The number of Native American scientists in the areas of biological sciences, bioinformatics, computational biology, genetics and data science is still low compared to other underrepresented minorities (URMs). Two areas that influence these numbers are the history of the assimilation of Native Americans in Western education and encounters of Western researchers' misuse of blood samples and genetic data. With this in mind, there are creative ways for institutions to create a positive space for engaging Native populations to enter the genomics field. Tribes are aware of scientific and technological advances and do not want to be left out of the discussion. In training the next generation of Native researchers, aligning cultural ideologies of Native people (i.e. natural resources and personal well being) to Western concepts in genomics is a big task. Here, discussion points on creating a positive space and engaging with Native communities will be presented.

W561: IWGSC - Standards and Protocols

Shaping Wheat for the Future: Leveraging the Wheat genome sequence in Crop Efficiency Research and Breeding

John Jacobs, Bayer CropScience NV, Gent, Belgium

Wheat is one of the three major staple crops worldwide, and the most widely grown crop on every continent except Antarctica. Despite increasing demands, wheat yield is stagnating and investments in wheat improvement are not keeping pace with other major crops, due to limited profitability of current business models.

Bayer believes this trend can be turned around with new technologies and wants to become a global leader in the Wheat Seeds & Traits Business. Since 2010 we are developing a strong R&D pipeline, encompassing both Breeding and Trait Research. The foundation of our wheat program is a broad germplasm base with local variety development in major wheat markets, a hybrid wheat platform and a portfolio of value adding traits, with a focus on agronomic performance.

The wheat genome sequence represents a key strategic resource for Bayer wheat research and breeding programs, ranging from the development and anchoring of molecular markers to support marker-assisted breeding, to the fine mapping, cloning and molecular characterization of genes underlying crop efficiency and hybrid performance.

At Bayer, we are convinced that a complete, high-quality reference genome sequence of bread wheat will be an invaluable resource to efficiently reach these objectives and the foundation for advancing public and private wheat research programs. For these reasons, Bayer CropScience is a strong supporter of the IWGSC since 2011, through both its contribution to the activities and strategy of the consortium, as well as the sponsoring of specific projects to accelerate the completion of key resources.

W562: IWGSC - Standards and Protocols

Sequencing and Analyses of Chromosome 1B

Frédéric Choulet[1], Hélène Rimbert[1], Ambre-Aurore Josselin[1], Benoît Darrier[1], Pierre Sourdille[1], François Balfourier[1], Valerie Barbe[2], Adriana Alberti[2], Karine Labadie[2], Sophie Mangenot[2], Arnaud Couloux[2], Patrick Wincker[2], Arnaud Bellec[3], Hélène Bergès[3], Michael Alaux[4], Thomas Letellier[4], Hadi Quesneville[4], Zeev M. Frenkel[5], Tzion Fahima[5], Abraham B. Korol[5] and Etienne Paux[1], (1)INRA GDEC, Clermont-Ferrand, France, (2)CEA - Genoscope, Evry, France, (3)INRA - CNRGV, Castanet Tolosan, France, (4)INRA - URGI, Versailles, France, (5)Institute of Evolution, University of Haifa, Haifa, Israel

We are currently producing a reference sequence of the bread wheat chromosome 1B (535 Mb) by sequencing 6023 BAC clones comprising the minimal tilling path of the physical maps of both short and long arms. BACs were sequenced by pools using a combination of Illumina MiSeq paired-end reads and HiSeq2500 5 kb mate pairs. First round of assembly yielded 7131 scaffolds representing 534 Mb. Report on the improvement of this assembly, the construction of a pseudomolecule and the comparison with the previous sequencing and assembly of chromosome 3B will be presented.

W563: IWGSC - Standards and Protocols

Sequencing and Assembling a Pseudomolecule for the Wheat Chromosome 6B

Hiroyuki Kanamori[1], Kanako Kurita[1], Harumi Sasaki[1], Satomi Mori[1], Satoshi Katagiri[1], Hiroko Fujisawa[1], Wataru Karasawa[1], Fuminori Kobayashi[1], Tsuyoshi Tanaka[1], Michihiko Shimomura[2], Nobukazu Namiki[2], Hiroshi Ikawa[2], Takashi Matsumoto[1], Yuichi Katayose[1], Jianzhong Wu[1] and Hirokazu Handa[1], (1)National Institute of Agrobiological Sciences, Tsukuba, Ibaraki, Japan, (2)Mitsubishi Space Software Co.,Ltd., Tsukuba, Ibaraki, Japan
The International Wheat Genome Sequencing Consortium (IWGSC) makes effort to obtain a high quality reference genome sequence of the bread wheat (cv. Chinese Spring), which is anchored to the genetic and physical maps. As a part of this international effort, we focus on sequencing of both (short and long) arms of chromosome 6B utilizing a minimum tiling path (MTP) BAC-by-BAC sequencing strategy.
A total of 7,573 MTP BAC clones were selected based on the physical mapping and then sequenced individually using 96-tag sequencing system with GS FLX (Roche). We found that up to 87% of these clones were successfully sequenced. For BAC clones with the successes on draft sequencing, we checked their sequence overlapping with neighboring BAC clones on the physical map, and it was confirmed that more than 95% of these BAC clones were truly overlapped. Two kinds of mate-pair libraries (3~4 kb, 8~11 kb) using pooled BAC clones was constructed and sequenced with MiSeq (24 tags/run, Illumina) to make a map-based and high-quality pseudomolecule of the wheat chromosome 6B. The assemblies of draft sequences from the BAC clones confirmed their overlapping were manually ordered using the output of SSPACE, which resulted in the first version of wheat chromosome 6B pseudomolecule (648.30 Mb). The number of scaffolds (> 10kb) was 2,935 and the size of the largest scaffold was about 3.1 Mb.
This work was supported by a grant from the Ministry of Agriculture, Forestry, and Fisheries of Japan. (Genomics-based Technology for Agricultural Improvement, NGB1003)

W564: IWGSC - Standards and Protocols

Never Eat Shredded Wheat – Navigating through Reference and Survey Wheat Assemblies Using Diverse Mate-Pair Libraries

David J. F. Konkin, National Research Council of Canada, Saskatoon, SK, Canada
David J. F. Konkin[1], Yifang Tan[1], Ron MacLachlan[2], Kevin Koh[1], Carling Clarke[1], Janet Condie[1], Jennifer Ens[2], Krysta Wiebe[2], Curtis Pozniak[2] and Andrew G. Sharpe[1]
[1]National Research Council Canada, Aquatic and Crop Resource Development, 110 Gymnasium Place, Saskatoon, Saskatchewan, Canada, S7N 0W9
[2]University of Saskatchewan, College of Agriculture, Department of Plant Science, 51 Campus Drive, Saskatoon, Saskatchewan, Canada, S7N 5A8
Long range sequence information is needed during genome assembly to span repetitive sequences and areas with limited sequencing depth. To improve existing wheat survey assemblies, an assembly of chromosome 7EL from *Thinopyrum elongatum*and to facilitate the assembly of reference sequences, we created 17 long range paired end libraries from a Chinese Spring + 7EL addition line using Illumina Nextera and Lucigen NxSeq approaches. We modified the Nextera protocol to create in parallel a large number of diverse libraries with tight size distributions that collectively span a large range (1 to 20 kb). We preferentially sequenced the most diverse libraries to greater depth using an Illumina Hiseq 2500 operating in rapid mode (2x150 bp). In total 584 Gbp of raw sequence data was acquired yielding 185 Gbp of useful processed sequence. Here I report the how we have used these sequences to improve to the IWGSC Chinese spring survey sequence, the whole genome shotgun assembly of synthetic hexaploid wheat W7984 and in the reference assembly of chromosome 1A.

W565: IWGSC - Standards and Protocols

Integrating Whole Genome Datasets into a BAC by BAC Approach to Wheat Chromosome Sequencing

Matt Clark, The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland

W566: IWGSC - Standards and Protocols

Application of Single Molecule Sequencing to Facilitate *de novo* Sequencing of Wheat Chromosome Arm 7DL

Song Weining, Northwest A&F University, Shaanxi, China
De novo sequencing of wheat chromosome 7DL has been carried out using the Illumina Hiseq2000 sequencing platform. This process is essentially based on a BAC-by-BAC approach with 4457 minimal tiling path (MTP) clones derived from a physical map of 7DL. Single molecule sequencing technology developed by Pacific Bio is a promising tool for the genomic analysis of complex genomes with high repetitive sequence content such as wheat. Initially Pac Bio sequencing was performed using P4-C2 chemistry with 4 BAC clones, combined with Illumina sequencing data. A contig per each BAC clone was obtained and the result was validated with PCR fragments short than 2 kb, although long-range validation is still pending. Later on, we pooled 47 BAC clones for Pac Bio sequencing using P6-C4 chemistry with ~10x coverage. With the assistance the Illumina sequencing data, a single contig per BAC could be obtained with almost half of the clones. It appears that single molecule sequencing technology is useful for the analysis of wheat BACs and further study with pooling size and software improvement could make this a valuable tool.

W567: IWGSC - Standards and Protocols

Construction of Pseudomolecules for the *Aegilops tauschii* Genome, the Wheat D Genome Progenitor

Tingting Zhu[1], Karin R. Deal[1], Shuyang Liu[2], Juan C. Rodriguez[2], Daniela Puiu[3], Geo Pertea[3], Steven L Salzberg[3], Aleksey Zimin[4], Frank M. You[5], Jan Dvorak[1] and **Ming-Cheng Luo**[1], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Department of Plant Sciences, University of California, Davis, CA, (3)Johns Hopkins University, School of Medicine, Baltimore, MD, (4)University of Maryland, College Park, MD, (5)Agriculture and Agri-Food Canada, Morden, MB, Canada
Aegilops tauschii is the donor of the D genome of hexaploid wheat (*Triticum aestivum*). Its large genome size and complexity impose a big challenge for sequencing and assembling the genome. Taking advantage of the available BAC-based physical map for the genome, MTP-based BAC pools were sequenced with Illumina's MiSeq platform. A comprehensive approach by integrating BAC pools sequences, whole genome shotgun Illumina sequences, PacBio sequences, and BioNano maps was used to assemble the genome and ultimately construct pseudomolecules

for each of the seven chromosomes. This work is part of the NSF-funded Project IOS-1238231 to generate a reference sequence for the genome of *Ae. tauschii* (http://aegilops.wheat.ucdavis.edu/ATGSP/).

W568: IWGSC - Standards and Protocols

Practical Aspects of BioNano Mapping of the Wheat Genome

Hana Simkova[1], Helena Stankova[1], Alex Hastie[2], Saki Chan[2], Jan Vrana[1], Zuzana Tulpova[1], Paul Visendi[3], Satomi Hayashi[3], Ming-Cheng Luo[4], Jacqueline Batley[5], David Edwards[6] and Jaroslav Dolezel[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)BioNano Genomics, San Diego, CA, (3)University of Queensland, Brisbane, Australia, (4)Department of Plant Sciences, University of California, Davis, Davis, CA, (5)University of Western Australia, Crawley, Australia, (6)University of Western Australia, Perth, Australia

Genome mapping in nanochannel arrays (IRYS platform), known as BioNano mapping, generates physical maps of a short sequence motif (recognition site of a nicking enzyme) along DNA stretches of hundreds to thousands kb in length. These can serve as a guide to validate and improve physical map and sequence assemblies.

Following the chromosome-based strategy of IWGSC, we proposed to couple BioNano mapping with chromosome flow sorting and construct BioNano maps for individual chromosomes/arms. A pilot experiment conducted on 7DS arm resulted in a high-resolution map, which, combined with BAC-by-BAC sequences of the 7DS arm, enabled validation, correction and anchoring of the BAC contig assembly, scaffolding sequences and sizing gaps and identification of large regions of tandem repeats. We have demonstrated that the BioNano map had a sufficient resolution to position and orient contigs consisting of as little as three BAC clones, a feature that can contribute to higher completeness of reference chromosome sequences. As such, the BioNano mapping provides a missing tool needed to complement the extant genomics tools to deliver high quality reference genome sequences and analyse structural genome variation. Subsequent projects on other group 7 chromosome arms confirmed that producing de novo BioNano maps for the whole wheat genome in chromosome-by-chromosome manner is feasible. Practical aspects, such as requirements on input sequence quality and current technology limitations will be discussed during the talk.

W569: IWGSC - Standards and Protocols

Towards a finished sequence for chromosome 7A: Building a high-quality pseudomolecule

Gabriel Keeble-Gagnere, Murdoch University, Perth, Australia

The latest progress in the assembly of the approximately 800Mb bread wheat chromosome 7A will be presented, including updates on the integration of mate-pair data and Bionano maps.

Particular challenges in the production of the whole-chromosome pseudomolecule will be discussed. In particular, an integrated approach for identifying and filtering cross-BAC-pool contamination, utilising mate-pair data, Bionano maps and the original BAC fingerprints will be described.

W570: IWGSC - Standards and Protocols

Hybrid assembly tools and strategies for a high quality wheat reference

Philippe Rigault, GYDLE, Québec, QC, Canada

W571: IWGSC - Standards and Protocols

IWGSC Sequence Repository: Moving Towards Tools to Facilitate Data Integration for the Reference Sequence of Wheat

Michael Alaux[1], Thomas Letellier[1], Françoise Alfama[1], Véronique Jamilloux[1], Jane Rogers[2], Frederic Choulet[3], Claire Guerche[1], Mikael Loaec[1], Raphael Flores[1], Célia Michotey[1], Anne-Francoise Adam-Blondon[1], Etienne Paux[4], Kellye Eversole[5] and Hadi Quesneville[1], (1)INRA - URGI, Versailles, France, (2)International Wheat Genome Sequencing Consortium, Cambridge, United Kingdom, (3)INRA UMR 1095 GDEC, Clermont-Ferrand, France, (4)INRA GDEC, Clermont-Ferrand, France, (5)Eversole Associates, Bethesda, MD

URGI is a genomics and bioinformatics research unit at INRA (French National institute for Agricultural Research), dedicated to plants and crop parasites. We develop and maintain a genomic and genetic Information System called GnpIS that manages multiple types of wheat data.

Under the umbrella of the IWGSC (International Wheat Genome Sequencing Consortium), we have set up a Sequence Repository on the Wheat@URGI website to store, browse and BLAST the data being generated by the wheat genome project: http://wheat-urgi.versailles.inra.fr/Seq-Repository.

The repository holds the wheat physical maps, the chromosome survey sequence data for the individual chromosomes of breadwheat, draft sequences for diploid and tetraploid wheats and provides browsable access to the BAC-based reference sequence for chromosome 3B, the first of the chromosomes to be completed by the consortium.

I will highlight the new features and data available in the Sequence Repository (e.g., new BLAST functionalities) and, in particular, present what we have done to address needs and concerns raised during the IWGSC S&P workshop last year. In addition, I will open the discussion about the future needs for tools to facilitate the integration of data to produce the reference sequence.

W572: IWGSC - Standards and Protocols

A Wheat Gene Expression Browser Implemented By the expVIP Customisable RNA-Seq Data Analysis and Visualisation Platform

Cristobal Uauy, John Innes Centre, Norwich, England, Philippa Borrill, John Innes Centre, Norwich, United Kingdom and Ricardo H. Ramirez-Gonzalez, The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland

The majority of RNA-seq expression studies in crop species remain underutilised and inaccessible due to the use of disparate transcriptome references and the lack of skills and resources to analyse and visualise this data. We have developed expVIP, an expression Visualisation and Integration Platform, which allows easy analysis of RNA-seq data combined with an intuitive and interactive interface. Users can analyse public

and user-specified datasets with minimal bioinformatics knowledge using the expVIP virtual machine. This generates a custom web browser to visualise, sort and filter the RNA-seq data and provides outputs for differential gene expression analysis. We demonstrate expVIP's suitability for polyploid crops and evaluate its performance in rice and wheat across a range of biologically-relevant scenarios. To exemplify its use in crop research we developed a flexible wheat expression browser (www.wheat-expression.com) which can be expanded with user-generated data in the local virtual machine environment. This open-access platform represents an opportunity to accelerate the rate of crop improvement by enabling the easy integration, visualisation and comparison of RNA-seq data across experiments.

W573: JBrowse, a Next Generation Genome Browser

JBrowse Installation and Configuration

Scott Cain, Ontario Institute for Cancer Research, Medina, OH

Tutorial Level

Beginner to Intermediate. Students should be comfortable performing simple command line tasks like moving files and running scripts.

Intended Audience

JBrowse is sufficiently easy to install (easier than GBrowse!) that a biologist can easily set up and configure a JBrowse server after the initial hurdles of learning about configuration options and file formats are overcome. This class is intended to help them over those hurdles.

Prerequisite Software and Conference PCs

Prerequisite software for JBrowse will be pre-installed on the conference PCs in the classroom area of the California Room. Participants using these PCs will be able to setup and configure JBrowse during the workshop.

After the workshop, a VirtualBox system image with JBrowse prerequisite software pre-installed will be made available on GMOD @ PAG page at GMOD.org. You can use this image to walk through the material presented at this workshop.

W574: Legumes

Nodulation Gene Networks in Legumes

Yupeng Li and Scott A. Jackson, University of Georgia, Athens, GA

With the accumulated -omics data sets, a major challenge has been to use these data to obtain a deeper understanding of the molecular mechanisms underlying complex traits. For example, root nodule symbiosis in legumes is one of the most productive nitrogen-fixing systems, fixing atmospheric nitrogen, and part of a sustainable agricultural cycle. Many genes and gene interactions involved in nodulation have been identified using traditional genetic and biochemical tools, but the complex nodule symbiosis process is far from fully understood. Here, three studies in gene network analysis have been done to help resolve this challenge in the post-genomics era. First, a nodulation gene network with 376 genes in *Medicago truncatula* was reconstructed using time-course transcriptome data during nodulation. Most of these genes are potentially novel nodulation-related genes. Their specific roles in nodulation have been predicted using module partition and functional analyses. Second, a new gene network reconstruction algorithm, weighted graphical lasso (wglasso), was developed to integrate multiple levels of -omics data. The algorithm significantly improved the accuracy of gene network reconstruction based on trancriptome data through the use other levels of gene interactions as prior knowledge. Third, a crowdsourcing platform was built for researchers to share and iteratively view nodulation gene network in *Lotus japonicus*. The comprehensive and accurate information for nodulation genes and gene interactions in the platform can be integrated into wglasso for better gene network reconstruction. Together, these products constitute a system that can be used to iteratively improve the prediction of nodulation gene networks. The system can be applied for other complex biological systems in order to quickly and systematically discover and understand molecular mechanisms underlying complex traits.

W575: Legumes

Expression Profiling of Iron Deficiency Chlorosis (IDC) in Soybean (*Glycine max*): Similarities and Differences Between Low Iron Supply and Alkaline Stress

Brian M Waters, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE

Alkaline soils comprise 30% of the earth and have low plant-available iron concentration, and can cause iron deficiency chlorosis (IDC). In the North-Central U.S., IDC causes soybean yield losses of $260 million annually. However, it is not known whether molecular responses to IDC are equivalent to those resulting from low iron supply. IDC tolerant and sensitive soybean lines from recurrent selection on high-pH soils in Nebraska provide a contrast to identify specific factors associated with IDC. We compared gene expression under alkaline and low iron conditions by sequencing RNA from roots of IDC tolerant and sensitive soybean lines grown hydroponically. We found both substantial overlap and substantial differences in differentially expressed genes when comparing iron deficiency and alkaline stress. Classical iron uptake genes were upregulated by both Fe deficiency and alkaline stress, including ferric-chelate reductase and ferrous transporter genes, however, their gene products did not function well at alkaline pH. In addition, genes in the phenylpropanoid synthesis pathway were upregulated in both alkaline and low Fe conditions. These genes lead to the production of fluorescent root exudate compounds, such as coumarins. Fluorescence of nutrient solution increased with alkaline treatment, and was higher in the IDC tolerant line. We hypothesize that root coumarin exudates become essential at alkaline pH where the classical iron uptake system does not function well. We are performing metabolomic profiling to identify these compounds. This work could result in new strategies to screen for IDC tolerance, and provide breeding targets to improve crop alkaline stress tolerance.

W576: Legumes

What's New in the Legume Information System and the Federated Legume Database Initiative

Jacqueline D. Campbell[1], Sudhansu Dash[2], Ethalinda Cannon[1], Alan M. Cleary[3], Wei Huang[1], Scott R. Kalberer[4], Alex G. Rice[2], Jugpreet Singh[5], Pooja E. Umale[2], Nathan T. Weeks[4], Andrew Wilkey[1], Christopher D. Town[6], Jeremy D. DeBarry[7], David Fernandez-Baca[1], Andrew D. Farmer[2] and Steven B. Cannon[4], (1)Iowa State University, Ames, IA, (2)National Center for Genome Resources (NCGR), Santa Fe, NM, (3)Montana State University, Bozeman, MT, (4)USDA-ARS-CICGRU, Ames, IA, (5)ORISE Fellow, USDA-ARS-CICGRU, Ames, IA, (6)J. Craig Venter Institute, Rockville, MD, (7)University of Arizona, Tucson, AZ

Over the last three years, the Legume Information System (http://legumeinfo.org ; LIS) has been redesigned to interoperate with additional websites serving specific legume research communities and to integrate diverse datasets across the crop and model legumes. To integrate datasets, LIS provides genome and map viewers, synteny mappings among all sequenced legume species. Additionally, LIS provides a set of gene families, allowing researchers and breeders to traverse across orthologous, homeologous and paralogous genes and genomic regions among the legumes – both at LIS and other websites. LIS partners with other data resources to help integrate data not housed at LIS (e.g. SoyBase, Phytozome, PeanutBase and others). We will describe how these resources can assist breeders and molecular geneticists in better understanding the physiological and molecular processes of legumes. We will also describe a NSF-funded project nicknamed 'Legume Federation' (http://legumefederation.org), highlighting common data formats, data collection templates, and protocols for continuing to build and extend resources relevant to enabling a "legume-fed" world.

W577: Legumes

Common Symbiotic Signaling Pathway: Filling the Gaps to Understand the Early Events of the Symbiotic Interaction Between Legumes and Rhizobia

Oswaldo Valdes-Lopez[1], Judith Salas-Oropeza[2], Socorro Sánchez-Correa[3], Carolina Isidra-Arellano[3], Maritza Zavaleta-Pastor[3], Pierre-Marc Delaux[4] and Jean-Michel Ané[5], (1)FES-Iztacala, National University of Mexico (UNAM), Tlalnepantla, Mexico, Mexico, (2)FES Iztacala UNA;, Tlalnepantla, Mexico, (3)FES Iztacala UNAM, Tlalnepantla, Mexico, (4)Laboratoire de Recherche en Sciences Végétales, Université de Tou- louse, Université Paul Sabatier, Tolouse, France, (5)Department of Agronomy - University of Wisconsin Madison, Madison, WI

Legumes have the ability to symbiotically interact with both arbuscular mycorrhiza (AM) fungi and nitroxigen-fixing rhizobia. Different studies have demonstrated that the early events of these symbioses are controlled by a set of genes belonging to the Common Symbiotic Pathway (CSP). Different attempts have been made to identify new regulators of the CSP. However, most of these studies were performed in late stages of these symbioses. Recent phosphoproteomic and transcriptomic analyses reveal that there are different proteins and genes whose phosphorylation or expression level are significantly modified after 1h treatment with LCOs and NFs. These data suggest the existence of potential yet-uncharacterized regulators. Thus, our objective is to identify and characterize new members of the CSP. To achieve this goal we performed a data mining analysis by using phosphoproteomic and transcriptomic data from *Medicago truncatula* roots treated with LCO or NFs. For this analysis we used three criteria: 1) increase in the expression or phosphorylation level by 1h treatment with NFs or LCOs treatments, 2) the candidate gene must be expressed in root or nodules only, and 3) the candidate gene must be absent in the *A. thaliana* genome. Currently, we have identified five candidates genes that fit with the three criteria. Here, we present evidence that support the role of one of the candidates in the establishment of the symbiosis with both AM and rhizobia.

W578: Legumes

The Genome Sequences of Cultivated Peanut's Diploid Ancestors

David Bertioli, University of Georgia, Athens, GA

Cultivated peanut (*Arachis hypogaea*) is an allotetraploid with closely related subgenomes of total size ~2.7 Gb (2n = 4x = 40; genome type AABB). This makes its genome assembly very challenging. Therefore as first step to understanding the genome of cultivated peanut, its diploid ancestors *A. duranensis* and *A. ipaënsis*, which contributed the A- and B-subgenomes respectively, were sequenced. We show that the sum of the diploid genomes is a good proxy for the cultivated peanut genome. However, there are distinct signs that the cultivated peanut genome is evolving by genetic exchange between the subgenomes. Based on remarkably high DNA identity of *A. ipaënsis* and the B-subgenome of cultivated peanut, and biogeographical evidence, we conclude that *A. ipaënsis* may be a direct descendant of the same population that contributed the B-subgenome to cultivated peanut.

W579: Legumes

Genomic and Geographic Diversity in Common Bean Based on SNP and GIS Information

Paul Gepts, University of California, Davis, CA

W580: Maize

The Implications of Ancient Maize Selection and Demography for Future Improvement

Tim Beissinger, USDA-ARS, University of Missouri, Columbia, MO

Maize domestication involved a population bottleneck followed by rapid expansion, both of which left a pronounced mark on genetic diversity across the genome. We studied the consequences of this demographic history using whole genome sequencing data from 23 landrace maize and 13 teosinte individuals. We estimate that at the onset of the bottleneck, maize population size was reduced to approximately 5% that of teosinte, but that subsequently maize has expanded to be several-fold larger than teosinte. We observe that purifying selection, as opposed to hard sweeps on new mutations, has been the primary force driving maize evolution. Importantly, maize's recent expansion has led to an increased efficiency of purifying selection at purging deleterious alleles from the gene pool. Based on these observations, we predict that polygenic selection may have also played an important role in shaping the genome of modern maize. To investigate the importance of polygenic selection, we describe a simple statistic, applicable to multi-generation breeding or experimental populations, to test for selection on polygenic traits that may not be significant based on genome-wide association studies or selection mapping. We demonstrate the applicability of this test using the Wisconsin Quality Synthetic breeding population.

W581: Maize

Leveraging the Diversity and Genomic Resources of Maize to Investigate Photosynthesis and Transpiration

Anthony J. Studer, University of Illinois, Urbana, IL

The vast genetic resources available in maize make it an ideal system for studying complex traits. However, the diversity of maize has been underutilized as a tool for studying photosynthesis. One important aspect of leaf photosynthesis is the relationship between CO_2 uptake and

transpirational water loss. Even a modest reduction in transpiration, while maintaining rates of photosynthesis, would constitute substantial water savings for maize cultivated under both irrigated and unirrigated conditions. These water savings are even more critical under water stress conditions, which are predicted to be more frequent and severe in the future (Cook et al., 2015 *Science Advances*: e1400082). Our research focuses on identifying the genetic control of leaf traits that alter photosynthesis and transpiration. The ratio of stable carbon isotopes (^{13}C and ^{12}C, ratio abbreviated $\delta^{13}C$) in leaf tissue is affected by the enzymes utilized by the photosynthetic pathway and by the plants water-use efficiency. We have identified significant variation in existing mapping populations for leaf $\delta^{13}C$ that is robust across environments, making it an excellent trait for genetic analysis and an entry point for subsequent detailed physiological characterization. Quantitative genetic approaches are being used to map variation in leaf $\delta^{13}C$ and link it to leaf morphology. These findings will be incorporated into elite germplasm and tested under agronomically relevant conditions.

W582: Maize

Auxin Signaling in Maize Inflorescences

Andrea Gallavotti, Waksman Institute, Rutgers University, Piscataway, NJ

In plants, small groups of pluripotent stem cells called axillary meristems are required for the formation of the branches and flowers that eventually establish plant and inflorescence architecture. In maize inflorescences, reproductive axillary meristems initiate at the axils of suppressed bract primordia at the flanks of inflorescence meristems. To ensure their proper formation, boundary regions need to be specified. We have identified two maize genes, *BARREN INFLORESCENCE1* and *BARREN INFLORESCENCE4* (*BIF1* and *BIF4*) that regulate the early steps required for inflorescence formation. *BIF1* and *BIF4* encode Aux/IAA proteins, key components of the auxin hormone signaling pathway that is essential for organogenesis, and regulate both suppressed bract and axillary meristem initiation in inflorescences. BIF1 and BIF4 are integral to auxin signaling modules that regulate the expression of the bHLH transcriptional regulator BARREN STALK1 (BA1) that is necessary for axillary meristem formation and shows a striking boundary expression pattern. These findings suggest that auxin signaling directly controls boundary domains during axillary meristem formation.

W583: Maize

Introducing Elements of High Performance Computing and Two-Way Epistasis into Genome-Wide Association Studies with Stepwise Model Selection

Alexander E. Lipka, University of Illinois, Urbana, IL

Although the genome-wide association study (GWAS) has emerged as a predominant approach for quantifying associations between genome-wide marker sets and many important traits, the statistical models that are typically used suffer from two major drawbacks. Specifically, these models test only one marker at a time and they only consider additive effects. As such, the typical GWAS model is not capable of quantifying the simultaneous contributions of multiple interacting genomic loci to phenotypic variation. To address these two drawbacks, we are developing a java-based program that augments a publicly available stepwise model selection module by also considering two-way epistatic interactions between marker pairs for inclusion into the final statistical model. With such a modification, it will become possible for researchers to use a GWAS model that can encapsulate many relationships between the genome and phenotype that are possible with various genetic architectures. Because the incorporation of epistatic effects introduces a substantial computational burden, we are also introducing high-performance computing approaches to this stepwise model selection program. In this talk, we use a subset of the US maize nested association mapping (NAM) population to demonstrate both the biological insight that is possible by incorporating two-way epistatic effects and the reduction of computational burden that can be realized through multithreading. We also discuss the direction we would like to take this stepwise model selection program so that it can accommodate the large-scale genomic data that are becoming increasingly available.

W584: Managing Crop Phenotype Data

Managing Phenotypic Data through Cassavabase with Fieldbook App

Alex C. Ogbonna[1,2], Agbona Afolabi[3], Guillaume J. Bauchet[1], Bryan Ellerbrock[1], Naama Menda[1], Chiedozie Egesi[2], Adeyemi Olojede[2], Ismail Rabbi[3], Peter Kulakow[3] and Lukas Mueller[1], (1)Boyce Thompson Institute for Plant Research, Ithaca, NY, (2)National Root Crops Research Institute (NRCRI), Umuahia, Nigeria, (3)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

Cassavabase (https://cassavabase.org/) is a central datastore for the Next Generation Cassava Project, aiming to leverage cassava genomic breeding and accelerate variety development in Africa (http://www.nextgencassava.org/). Cassavabase supports breeder's daily activities as a comprehensive information source on phenotypes and genotypes through Cornell's IGD LIMS system. As well, it provides tools for performing trait analysis, genomic selection and other tools such as maps, genome browser. Cassavabase is a globally accessible resource, and data can be contributed and viewed by any registered user.

Traditionally, much of the phenotypic information generated from a breeding program is: access restricted to users as stored in paper based field books or researchers' computer files. To efficiently address these issues on phenotypic data collection, Cassavabase developed a pipeline to collect and upload phenotypes.

Adopting the Fieldbook app (KSU website) and an Android tablet, Cassavabase offers an efficient tookit to address field data collection. The fieldbook app provides a user friendly interface, large autonomy for data collection, flexible file transfer and has little glare while being used in the field for data recording. However, this presentation forms part of the dialogue to address the many challenges that threatens the continues adoption of the Android tablet for field data recording. Cassavabase enables large data storage (7 millions phenotypic datapoints currently stored) and uses of a cassava specific trait ontology dictionary, listing users' phenotypic trait description and measurement method. Altogether this toolkit facilitates Field to Office data transfer, making information quickly and accurately available for further analysis (I.e: descriptive statistics, genomic selection)

In this workshop, we will highlight the practical aspects of field data collection and uploading using cassavabase with two use cases including field data collection and trait ontology management. We will highlight challenges encountered that can be addressed by ongoing dialogue between field teams and developers of the fieldbook app and Cassavabase.

W585: Managing Crop Phenotype Data

Genomes To Fields (G2F)

Carolyn Dill, Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA

Genomes to Fields (G2F) is an umbrella initiative to support translation of maize genomic information for the benefit of growers, consumers and society. This public-private partnership is building on publicly funded corn genome sequencing projects to develop approaches to understand the functions of corn genes and specific alleles across environments. Ultimately this information will enable prediction of the phenotypes of corn plants in diverse environments. There are many dimensions to the over-arching goal of understanding genotype-by-environment (GxE) interactions, including which genes impact which traits and trait components, how genes interact among themselves (GxG), the relevance of specific genes under different growing conditions, and how these genes influence plant growth during various stages of development. This initiative promotes projects that advance integrated research and technologies, combining fields such as genetics, genomics, plant physiology, agronomy, climatology and crop modeling, with computation and informatics, statistics and engineering. G2F's GxE subgroup - which is comprised mainly of maize breeders - aims to assess environmental effects on an extensive collection of inbreds and hybrids at environmentally diverse locations.

My lab group is working to identify, develop, and deploy solutions to manage genotype, environment, and phenotype data. For the GxE subgroup we support current data management needs simultaneous with planning for the deployment of improved information platforms. Current platform components include resources developed by the iPlant Collaborative and the Integrated Breeding Platform's Breeding Management System (BMS). Data types, hard problems, lessons learned, and likely partners for future collaborative development will be presented and discussed.

W586: Managing Crop Phenotype Data

GnpIS-Ephesis, the Phenotypic Data Integration Platform for Inra Networks Experimental Data – Data Discovery and Dataset Building Use Cases

Cyril Pommier[1], Michael Alaux[1], Thomas Letellier[1], Célia Michotey[1], Guillaume Cornut[1], Aristide Lebreton[1], Mathieu Labernadière[1], Mathide Lainé[1], Elizabeth Arnaud[2], Anne-Francoise Adam-Blondon[1] and Hadi Quesneville[1], (1)INRA - URGI, Versailles, France, (2)Bioversity International, Montpellier Cedex 5, France

Phenotype data are collected in trials conducted by experimental facilities including multilocal field networks and high throughput phenotyping facilities in controlled environments or fields. A given germplasm panel can therefore have been phenotyped in very different conditions and using very different protocols. As a result, a collection of phenotype datasets is usually highly heterogeneous and hard to integrate.

GnpIS is an integrative information system dedicated to plant and their pathogens. The integration of heterogeneous phenotypic datasets implies identifying common pivot resources like germplasm, observation variables following the Cropontology model, experimental locations and years. GnpIS allows performing data discovery on those data, which can lead to datasets building through the GnpIS-Ephesis application.

GnpIS-Ephesis allows creating datasets to study the relations between yield, including its components, stress and disease tolerance from public provided by the INRA Wheat Network Phenotypic. It includes fifteen years of observations on eleven experimental sites. Those study datasets can be narrowed from a maximum number of observations on several years and locations to a minimal dataset on comparable locations/years pairs to reduce the environmental variability. This variability can be evaluated thanks to reference germplasms.

Quercus, Populus or Vitis public data available in GnpIS-Ephesis allows to study adaptation to climate change. For instance, a dataset including phenology variables like budbreak or flowering can be extracted and used as input for statistical analysis tools or model to evaluate adaptability of several hundreds of vitis variety.

Aknowledgment

Duchêne E, Lacombe T, Oury FX, Charmet G, Gauffretau A

W587: Managing Crop Phenotype Data

Android Apps for Plant Breeding and Genetics #phenoApps

Trevor W. Rife and Jesse Poland, Kansas State University, Manhattan, KS

Plant breeding and genetics research is an inherently data-driven enterprise. Typical experiments and breeding nurseries can contain thousands of unique entries and programs will often evaluate tens of thousands of plots each year. To operate efficiently on this scale, electronic data management becomes essential. Many research programs, however, continue to operate by scribing and transcribing massive amounts of data on paper field books. While effective, this form of data management places heavy burdens on human resources, decreases data integrity, and limits future utilization of data and the ability to expand the breeding program. To help address these constraints, we have developed Field Book, an open-source application for electronic data capture that runs on consumer-grade Android devices. We attempt to decrease both technological and cost barriers that hinder adoption of electronic data management in breeding programs by focusing on a simple, stand-alone application with an intuitive and customized interface. Integrating low-cost hardware into breeding programs will provide an inexpensive platform for additional apps to be developed and distributed, providing accessible solutions for breeding programs around the world. Transformational capacity in electronic data collection and management will be essential to realize a contemporary green revolution.

W588: Managing Crop Phenotype Data

Triticeae Toolbox (T3)

Kevin P. Smith, Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN

The Triticeae Toolbox (T3), formerly the Horduem Toolbox, was initially created to store trait and genotype data from small grains breeding programs to facilitate association mapping. Since its inception, T3 has evolved to become an integral component of breeding programs providing more than a data repository. T3 facilitates data sharing among researchers providing both data integrity and the ability to select data sets using a wide range of criteria for a variety of purposes. In addition, there are a growing number of data visualization (trait distributions, PCA plots, haplotypes, and pedigrees) and analysis tools (association mapping, genomic prediction, blast searches, and selection indices) that can be run

directly from the database interface making exploratory analysis fast and simple. Together these tools and the growing database enable collaborative research projects by allowing multiple users to access and interact with data to accomplish multi-stage tasks. To illustrate this, we will show how T3 was used to initiate a new two-row breeding program at the University of Minnesota.

W589: Managing Crop Phenotype Data

Breeding Management System (BMS): A User Perspective

M. Cinta Romay, Institute for Genomic Diversity, Cornell University, Ithaca, NY

While getting information about the genome is becoming cheaper, the cost of obtaining phenotypic information has maintained or even increased. The number of genomic related hypothesis that researchers want to test is becoming much bigger than the access to phenotypic data that each of those researchers can obtain and combine. In order to overcome this issue, phenotypes from thousands of plots need to be easy to access and well organized. The breeding management system (BMS) can help consolidate, and share large amounts of phenotypic data. With an intuitive and user friendly interface that can be installed locally or in a server, the software provides an easy platform to organize different experiments, manage seed inventory, and upload, download, and share phenotypic data. The program integrates well with electronic data capture systems like fieldbook and has a crop ontology feature that allows the users to clearly choose and define the traits being measured. However, there is still room from improvement and tools for communication with the developers and community development are being put in place to help implement new features that users may need. The scope of this presentation is to show the advantages and disadvantages of this software from the perspective of a quantitative genetics lab, and how the Buckler lab has been able to modify some features to adapt it to its needs in close collaboration with the system developers.

W590: Mango genomics

Accurate Assembly and Phasing of Heterozygote Genomes

Kobi Baruch, NRGENE LTD., Ness-Ziona, Israel

Dr. Kobi Baruch, NRGene

Accurate denovo assembly of heterozygote genomes is highly complex. To cope with this challenge, NRGene has developed its unique assembly solution – DeNovoMAGIC.

The talk will describe DeNovoMAGIC features and advantages versus other approaches and share initial results of the assembly of the heterozygote mango genome.

Title:**Accurate Assembly and Phasing of Heterozygote Genomes**

Submitter's E-mail Address: kobi@nrgene.com

W591: Mango genomics

The Queensland Mango Genomics Initiative

Natalie Dillon[1], **David Innes**[2], Roger Broadley[3], Rajeev K Varshney[4] and Ian Bally[1], (1)Queensland Department of Agriculture and Fisheries, Mareeba, Australia, (2)Queensland Department of Agriculture and Fisheries, Dutton Park, Australia, (3)Department of Agriculture and Fisheries, Nambour, Australia, (4)ICRISAT, Hyderabad, India

Mango is an important industry for Queensland, Australia, with an annual value exceeding $80 million. The Kensington Pride cultivar, prized by consumers for desirable taste and colour characteristics, commands 60% of the domestic market though this market share has declined in recent years as new varieties, such as Calypso™, get established with consumers. In 2005, the Queensland Government's Department of Agriculture and Fisheries commenced the Mango Genomics Initiative. This project brought together multidisciplinary teams of breeders, pathologists, sensory scientists, flavour chemists and molecular biologists to develop a suite of tools and inter-related data sets to support the accelerated development of new commercial mango varieties. An overview of the Mango Genomics Initiative will be presented here culminating in the generation of a draft Kensington Pride mango genome sequence.

W592: Mango genomics

A Genetic Map and Germplasm Diversity Estimation of *Mangifera indica* (Mango) with SNPs

David N. Kuhn[1], Yuval Cohen[2], Amir Sherman[3], Ron Ophir[4], Amy Groh[5], Jordon Rahaman[5], Paola Sanchez[5], Natalie Dillon[6], Ian Bally[6] and David Innes[7], (1)USDA ARS SHRS, Miami, FL, (2)Volcani Research Center, Bet Dagan, Israel, (3)ARO, Bet Dagan, Israel, (4)Agriculture Research Organization, Volcani Center, Bet Dagan, Israel, (5)USDA-ARS, Miami, FL, (6)Queensland Department of Agriculture and Fisheries, Mareeba, Australia, (7)Queensland Department of Agriculture and Fisheries, Dutton Park, Australia

Mango (*Mangifera indica)* is often referred to as the “King of Fruits”. As the first steps in developing a mango genomics project, we genotyped 582 individuals comprising six mapping populations with 1054 SNP markers. The resulting consensus map had 20 linkage groups defined by 726 SNP markers with 67 SNPs as the maximum number for one linkage group and 20 SNPs the minimum number. Although mango has been described as an allotetraploid with 40 chromosomes (n = 10), all SNPs showed disomic inheritance and Mendelian segregation patterns of a diploid with n = 20. SNP markers (384) that were evenly distributed across the mango genome as defined by map position were used to genotype ~800 individuals including other *Mangifera* species from several germplasm collections in a first round of germplasm genetic diversity estimation. Clustering and STRUCTURE analysis of the germplasm data will be presented.

W593: Mango genomics

Mango Genetics

Amir Sherman, ARO, Bet Dagan, Israel

Mango is one of the most important fruit crops in tropical and subtropical regions. We assembled the transcriptomes of two mango cultivars, Keitt and Tommy-Atkins and use them as a basis for our genetic and genomic analyses. Genetic variation was used to analyze the structure of our germplasm collection and to explore parentage relations of our accessions from our breeding programs. As mango crosses using hand

pollinations are complicated and very inefficient, we established a post pollination assays for identification of paternal parents generated through open pollination based on unique SNPs signatures. Using this tool and high throughput genotyping platform we develop 7 different F1 populations that share the same accession as their maternal parent. These will form a unique resource for mango breeding. Based on the mango transcriptome we isolated a set of genes involved in pathways controlling fruit peel color. Combining expression, phenotyping and genotyping, we investigating the molecular basis for mango peel color. These approaches will facilitate the identification of markers assisted breeding for mango peel color.

W594: Mango genomics

Mango Transcriptomes

Maria A. Islas-Osuna, Centro de Investigación en Alimentación y Desarrollo, Hermosillo, Mexico

The transcriptomes of mango mesocarp cv. “Kent” (GenBank PRJNA258477) and “Ataulfo” (GenBank PRJNA286253) were obtained using the Illumina sequencing plataform. A total of 52,948 putative genes were obtained for “Kent” and 40,094 for “Ataulfo”, and their gene products were assigned to 327 and 341 biochemical pathways, respectively. Differential expression between mature green and ripe fruit was done for “Kent” and “Ataulfo”. Additionally, the effect of the hot water quarantine treatment was studied in “Ataulfo” mango by RNA-seq differential expression.

Starch and sucrose metabolism, fatty acid metabolism and carbohydrate catabolic process were up-regulated and enriched during ripening and by the hot water treatment. Several gene families associated to cell wall degradation were identified and compared within family members. The expression of 26 genes involved in pectin degradation were evaluated by qRT-PCR and their levels were similar to those obtained by RNA-seq. Enzymes up-regulated due to ripening and hot water treatment will be discussed. Also, coding sequences for MADS-box transcription factors that regulate development and fruit ripening were found in the mango transcriptomes and will be presented.

Small heat shock proteins and heat shock factors were induced due to the heat stress in mango fruit and will be discussed. These mango transcriptomes have been very helpful to identify genes for expression studies related not only to ripening but also to the responses to stress.

W595: Mango genomics

Assembly of highly heterozygous mango (*Mangifera indica* cv. Amrapali) genome using PacBio long sequence reads

Nagendra K. Singh[1], Ajay K. Mahato[1], Pawan K. Jayaswal[1], Akshay Singh[1], Sangeeta Singh[1], Vandna Rai[1], Amitha Mithra S. V.[1], Kishor Gaikwad[1], Anand K. Singh[2], Nimisha Sharma[2], Manish Srivastava[2], Jai Prakash[2], Usha Kalidindi[2], S. K. Singh[2], Kasim Khan[3], Rupesh K. Mishra[3], Shailendra Rajan[3], Anju Bajpai[3], B.S. Sandhya[4], Puttaraju Nischita[4], K. V. Ravishankar[4], M.R. Dinesh[4], Neeraj Kumar[5], Sarika Jaiswal[5], Mir A. Iquebal[5], Dinesh Kumar[5], Anil Rai[5] and Tilak R. Sharma[1], (1)ICAR-National Research Centre on Plant Biotechnology, New Delhi, India, (2)ICAR-Indian Agricultural Research Institute, New Delhi, India, (3)ICAR-Central Institute for Subtropical Horticulture, Lucknow, India, (4)ICAR-Indian Institute of Horticultural Research, Bengaluru, India, (5)ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Mango (*Mangifera indica* L.), a member of the family Anacardiaceae is an allotetraploid fruit tree with small genome size of 439 Mb (2n= 40). India is the largest producer of mango growing more than 1,000 varieties, where it is known as ‘king of fruits’ for its rich taste, flavor, color and diverse end-usage. To facilitate the assembly of highly heterozygous diploid genome of a popular mango variety ‘Amrapali’ we used PacBio-SMRT sequencing with long average read lengths and 70x genome coverage. *De novo* assembly using FALCON diploid assembler resulted in an assembly of 323 Mbp in 9,550 large contigs, with 73.2% genome coverage, largest contig size of 1.09 Mb and N50 of 98.3 Kb. Gene prediction using FGENESH revealed total 43,247 gene models with an average gene size of 894 bps in a range of 150 to12,102 bp. Comparison with the other published plant genomes showed highest similarity with *Citrus sinensis*. We identified 122,332 genomic SSR and developed 8,451 Type1 SSR and 835 HSSR markers with high level of detectable polymorphism. Further, we have identified 1.67 million high quality SNPs by ddRAD sequencing of 91 diverse mango varieties for which a database has been created and population structure of Indian mango varieties is determined. Using this data a single-copy gene based 50K SNP chip has been designed for genotyping using Affymetrix platform. These genomic resources will fast track the mango varietal improvement for higher productivity, disease resistance and superior end-use quality.

W596: Mutation Screening

DNA-Free Genome Editing in Plants with Preassembled CRISPR-Cas9 Ribonucleoproteins

Sunghwa Choe, Seoul National University, Seoul, South Korea

Site-directed mutagenesis has been widely employed in plants due to their crucial role in gene-function studies. To expedite societal use of genome-edited crops, we aimed to develop an alternative way of plant genome editing without introducing foreign DNA into cells. DNA-free methods may alleviate regulatory concerns related to genetically modified plants. To this end, we transfected preassembled complexes of purified Cas9 protein and guide RNA into plant protoplasts of Arabidopsis thaliana, tobacco, lettuce and rice and achieved targeted mutagenesis in regenerated plants at frequencies of up to 46%. The targeted sites contained germline-transmissible small insertions or deletions that are indistinguishable from naturally occurring genetic variation. Therefore, it should be better to treat the crop plants engineered with DNA-free genome editing methods as the varieties generated after conventional mutagenesis and even natural spontaneous mutations.

W597: Mutation Screening

CRISPR/Cas9-Mediated Viral Interference in Plants

Magdy Mahfouz, Biological and Environmental Sciences and Engineering Division, Thuwal, Saudi Arabia

The CRISPR/Cas9 system provides bacterial and archaeal species with molecular immunity against invading phages and conjugative plasmids. The CRISPR/Cas9 has been used for targeted genome editing in diverse eukaryotic species including plants. We investigated whether the CRISPR/Cas9 system could be used in plants to confer molecular immunity against DNA viruses. Indeed, our data show that the CRISPR/Cas9 system is portable to plants. CRISPR/Cas9 machinery including sgRNAs targeting the coding and non-coding sequences of *Tomato yellow leaf curl virus* (TYLCV) targeted the virus for degradation and introduced mutations at the target sequences. Using a single sgRNA targeting a

conserved sequence in Geminiviruses resulted in a simultaneous viral interference against multiple viruses. Taken together, our data establish the efficacy of the CRISPR/Cas9 system for viral interference in plants for addressing basic questions and for biotechnological applications.

W598: Mutation Screening

Analysis of an EMS Mutagenized Population of Wheat by Exome Capture Identifies Widespread Deletions

Andy L Phillips[1], Robert King[1], Ricardo H. Ramirez-Gonzalez[2], Paul C. Bailey[2], Ksenia V Krasileva[3], James Simmonds[4] and Cristobal Uauy[5], (1)Rothamsted Research, Harpenden, United Kingdom, (2)The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland, (3)The Genome Analysis Centre, The Sainsbury Laboratory, Norwich, United Kingdom, (4)John Innes Center, Norwich, United Kingdom, (5)John Innes Centre, Norwich, United Kingdom

Mutagenesis using ethyl methane sulphonate (EMS) is a common method for increasing available genetic variation in plants: the mutations created can be used for both forward and reverse genetics. The latter approach is particularly powerful in polyploid species such as wheat as genetic redundancy limits the use of forward genetics but also allows very high mutation rates. We and our collaborators have developed a high-throughput mutation screening approach in wheat using exome capture and next-generation sequencing. This has identified a very large number of point mutations in the majority of wheat genes within an EMS-mutagenized population of wheat cv. Cadenza that provides a publicly-available resource for functional genomic and crop improvement. In this work, we have analyzed the exome capture data for variation in read coverage at the scaffold and exon level and have identified widespread deletions, including the loss of whole chromosome and chromosome arms. Large homozygous deletions were observed to be far less abundant in the M_2 DNA than heterozygous deletions, indicating that large deletions may affect transmission of gametes or viability of offspring. Data will be presented on the presence of such deletions in subsequent generations of the population.

W599: Mutation Screening

The Sorghum *brown midrib* (*bmr*) Mutants: A Forward Genetics Approach to Lignin

Scott Sattler, USDA-ARS Grain, Forage and Bioenergy Unit, Lincoln, NE

The *brown midrib* (*bmr*) mutants of sorghum and other C4 grasses have a visible red to brown colored leaf midrib in contrast to green or white midribs observed in wild-type plants. This phenotype, more importantly, is associated with reduced lignin content and altered lignin composition, which have led to interest in using *bmr* mutants to improve cellulosic bioenergy conversion and increase forage digestibility for livestock. The sorghum *bmr* mutants were generated through chemical mutagenesis (EMS and DES), and, through allelism testing eight complementation groups were identified in sorghum. Three of the *bmr* loci (*bmr2*, *bmr6* and *bmr12*) have been identified and shown to encode enzymes in monolignol biosynthesis (4-coumarate-CoA ligase, 4CL; cinnamyl alcohol dehydrogenase, CAD; caffeic acid-O-methyltransferase, COMT), the pathway that synthesizes the subunits of lignin. The effects of these mutations on agronomic traits, lignin content, lignin composition and enzymatic saccharification were characterized. Nearly all of mutants had reduced lignin concentrations in their stover (biomass) relative to the wild-type. However, increased enzymatic saccharification of *bmr* mutant stover relative to wild-type stover was not universally observed. For the alleles of the three identified loci, the impact of the mutations of their respective transcript levels, protein abundance and enzymatic activity were also characterized. Overall, these molecular analyses indicate that while the majority of alleles characterized are likely the result of amorphic (null) mutations, a few of alleles characterized likely represent hypomorphic (weak) mutations. These *bmr* mutants represent breeding tools for manipulating biomass composition to enhance forage and feedstock quality.

W600: Mutation Screening

Designing Plants with Novel Traits Using Forward and Reverse Genetics- the BenchBio Company Perspective

Manash Chatterjee, NUIG, Ireland & BenchBio Pvt Ltd, Vapi, India

The challenge of the future will be the speed at which DNA sequence databases are analysed and candidate genes of agronomic importance selected and tested for function. The problem in this strategy is the lack of high throughput (HTP) tools to test the function of specific genes in a cellular context. Transgenic (GM) method is one route to validate the gene function and to also carry out crop improvement. However, this is not available in all crops. Furthermore, GM crops continue to face tremendous difficulties in public acceptance; the technology is expensive and also frequently limits ‘freedom to operate’ (FTO). Additionally, at present the GM method cannot keep pace with the speed at which candidate genes for important traits are being identified and prioritised for functional evaluation. At BenchBio, we are using an alternative tool; saturation mutagenesis in combination with forward and reverse genetics which offers an alternative way to improve crops without transgenics. Reverse genetics combines traditional chemical mutagenesis with sensitive molecular screenings to discover induced point mutations in genes controlling important traits whose sequence is known. Forward genetics combines identification of novel traits by phenotyping and cloning of genes afterwards. By combining these two methods new genotypes with potentially high agronomic value can be isolated and directly commercialized. We will present data to demonstrate that these methods can be used successfully to improve crops

W601: Mutation Screening

Identification of Rare Alleles in Soybean using TILLInG by Sequencing

Karen Hudson, USDA-ARS, West Lafayette, IN, Rima Thapa, Purdue University, WEST LAFAYETTE, IN and Katy Martin Rainey, Purdue University, West Lafayette, IN

One of the significant problems with establishing a TILLING resource for soybean is the paleopolyploid origin of soybean, which means that multiple highly similar sequences are present for each gene. This creates both technical complications for amplification and sequence analysis, as well as a practical problem that both highly similar copies may have to be mutated in order to obtain the desired effect of a gene knock-out. These factors make TILLING in soybean cost prohibitive. In this project we have pilot tested a method to survey the available diversity and select the most damaging mutations and identify the individuals that carry them by genotyping pools and subsequently individual plants. We have used this approach to identify mutations in genes that affect seed carbohydrate biosynthesis. The reverse-genetics approach will be compared to forward genetics approaches in soybean in which we have also utilized next generation sequencing to identify causative alleles.

W602: Mutation Screening

Tilling by Sequencing for Genome-Wide Mutation Discovery and Functional Genomics in *Camelina sativa*

Sateesh Kagale[1,2], Wayne E. Clarke[2], Lily Tang[2], Brittany Polley[2], Chushin Koh[1], Eric Johnson[3], Andrew G. Sharpe[4] and Isobel Parkin[2], (1)National Research Council Canada, Saskatoon, SK, Canada, (2)Agriculture and Agri-Food Canada, Saskatoon, SK, Canada, (3)Agriculture and Agri-Food Canada, Scott, SK, Canada, (4)National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada

TILLING (Targeting Induced Local Lesions in Genomes) is a non-transgenic reverse genetic approach that allows genome-wide discovery of induced point mutations in EMS-mutagenized populations. The development of genotyping-by-sequencing (GBS) to rapidly detect nucleotide variation at the whole genome level, in highly multiplexed sets of discrete individuals simultaneously has provided a transformative genetic profiling technique. We have successfully established a GBS-based approach for high-throughput screening for mutations in an EMS-mutagenized population of *Camelina sativa*. To assay the mutation rate and efficacy of the GBS approach, ~400 M_2 lines were grown, and genomic DNA and seeds were collected from ~200 individuals. Two separate DNA pools of 96 individuals, for a total of 192, were constructed by mixing equal amounts of genomic DNA. GBS analysis of these individuals utilized two methylation sensitive restriction enzymes (*Pst*I and *Msp*I), one with a six-base and the second with four-base recognition site, to effect genome reduction. GBS libraries were sequenced on Illumina HiSeq 2000 platform. A customized bioinformatics pipeline identified 6,057 unique candidate EMS SNPs of which 2,594 were exonic and the remaining aligned to non-exonic regions. The number of mutations in each line varied significantly potentially due to varying EMS penetration or lack of mutagenization in seeds. High proportions of SNPs were EMS induced as ~90% of the recovered mutations were C-to-T and G-to-A transitions, a hallmark feature of EMS mutagenesis. A detailed functional analysis and phenotypic characterization of recovered mutations will be presented. Additionally, potential applications of this effective approach of mutation discovery to other polyploid crops, such as Brassicas and wheat will be discussed.

W603: National Plant Genome Initiative Workshop - Challenges and Opportunities in High-Throughput Phenotyping and Data Management

Genomes to Fields: Translating Our Understanding of the Genome to Predictions of Performance in the Field

Nathan M. Springer, University of Minnesota, St. Paul, MN

Agriculture faces the enormous challenge of efficiently and sustainably producing a safe, dependable food supply for a growing world population. Meeting this challenge requires the development and management of crop varieties that will perform well in spite of increased weather variability. To address this challenge, an ambitious widescale plant phenotyping initiative is proposed which will expand our understanding of the interacting roles of crop genomes and crop environments (including weather and management practices) on crop performance. A critical part of the Genomes To Fields Initiative is fostering interdisciplinary research collaborations between crop scientists and engineers and computational scientists. The engineers will develop new technologies (e.g., robots and sensors) that will make high-throughput phenotyping possible. The computational scientists will develop new computational approaches to analyze the resulting very large (big data) datasets of phenotypes and environmental data and thereby facilitate an understanding of gene function and to address fundamental problems of agricultural productivity. The Initiative's interdisciplinary network of scientists and datasets will provide an excellent training platform for students studying agriculture, genetics and engineering. By improving our ability to predict crop performance in diverse environments, this initiative will enhance our ability to develop new varieties and to manage the effects of weather variability on crop production.

W604: National Plant Genome Initiative Workshop - Challenges and Opportunities in High-Throughput Phenotyping and Data Management

Genomic and Open-source Breeding Informatics Initiative (GOBII)

Edward S. Buckler[1], Kelly Robbins[2], Jean-Luc Jannink[2], Tobias Kretzschmar[3], Lukas Mueller[4], Yaw A. Nti-Addae[2], Michael S. Olsen[5], Mark E Sorrells[2], Qi Sun[6], Rajeev K Varshney[7] and Susan McCouch[2], (1)USDA-ARS-Cornell University, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)International Rice Research Institute, Los Banos, Philippines, (4)Boyce Thompson Institute for Plant Research, Ithaca, NY, (5)CIMMYT, Nairobi, Kenya, (6)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (7)ICRISAT, Hyderabad, India

Tremendous progress in that last decade has been made in developing marker systems that are less expensive than field trials, and statistical approaches for relating phenotype to genotype that have substantial accuracy. However, the bioinformatics of uniting high throughput markers, whole genome sequence, and phenotypes from around the world is challenging for all groups. The Genomic & Open-source Breeding Informatics Initiative (GOBII) is the first large-scale public-sector effort to enable systematic application of high-density genotypic information to the breeding of staple crops in the developing world. The project will develop and implement genomic data management systems to enhance the capacity of public-sector breeding programs to deliver increased rates of genetic gain in South Asia and Sub-Saharan Africa for rice, wheat, maize, sorghum, and chickpea. The genomic data management systems will include databases, analysis pipelines, and decision support tools for plant breeders. The open source nature of this project also provides opportunities for all sorts of other research groups to use this software stack in their own research.

W605: National Plant Genome Initiative Workshop - Challenges and Opportunities in High-Throughput Phenotyping and Data Management

Challenges and Opportunities in High-Throughput Field Phenotyping

Sindhuja Sankaran[1], Arron H. Carter[1], Lee A. Kalcsits[2], Jack K. Okamuro[3], David C. Slaughter[4], Helmut Kirchhoff[1] and Jesse Poland[5], (1)Washington State University, Pullman, WA, (2)WSU Tree Fruit Research and Extension Center, Wenatchee, WA, (3)USDA ARS, Beltsville, MD, (4)University of California, Davis, Davis, CA, (5)Kansas State University, Manhattan, KS

In last few years, versatile sensors and associated instrumentation with cross platform connectivity have provided opportunities to enhance our ability to rapidly phenotype in both controlled environment and field conditions. Several research programs across the U.S. are integrating such

advanced phenotyping tools into the breeding programs. Overall, developments to date are crop- and area-specific addressing local breeding program needs and often independent of each other. To gain better understanding and have national integration of such efforts, our team has organized a conference in November 2015. Tree fruit, grain, legume and other specialty crop breeders, plant physiologists, data management experts, and sensor and automation experts participated in this two day event to (i) identify current status, gaps, and future research needs for field-based high throughput phenotyping and data management, and (ii) integrate multiple efforts to develop a platform for resource sharing. This talk is aimed to brief the outcomes of this conference.

W606: National Plant Genome Initiative Workshop - Challenges and Opportunities in High-Throughput Phenotyping and Data Management

Title and abstract pending.

Todd C. Mockler, Donald Danforth Plant Science Center, Saint Louis, MO

W607: National Plant Genome Initiative Workshop - Challenges and Opportunities in High-Throughput Phenotyping and Data Management

Quantitative Iformation on Pants in Variable and Heterogeneous Environment - Integrated Approaches to Mechanistic, High-Throughput and Field Phenotyping

Ulrich Schurr, Forschungszentrum Jülich GmbH, Jülich, Germany

Plant phenotyping develops rapidly into a bottleneck for progress in basic and applied research. Lack of adequate solutions for quantitative analysis of plant architecture and function as well as their interaction with the dynamic and heterogeneous environment hampers progress in basic sciences as well as in breeding-related research. In recent years significant interdisciplinary approaches have been started to overcome this “phenotyping bottleneck”. Techniques were developed to quantify the dynamics and the heterogeneity of plant structure and function as well as of environmental cues. These mostly non-invasive technologies are developed and implemented into biological concepts that allow novel insights in the dynamic characteristics of plants above- as well as belowground. The technologies include high-resolution analysis for mechanistic understanding (like MRI and PET for structure, growth and activity of roots and shoots), the high-throughput approach for analysis of large numbers of genotypes and environmental conditions as well as field approaches, which are the reference to indicate the relevance. The talk will provide an overview on recent developments in technologies as well as conceptual approaches as the basis for a quantitative understanding of plant-environment-dynamics and its application for plant breeding and plant management. The talk will also present recent developments in infrastructure platforms that have been and will be established in Germany, in Europe and globally.

W608: NCBI Genome Resources

Simplified Access to SRA Genomic Data within GATK

Stephen Sherry, NIH/NLM/NCBI, Bethesda, MD

The NCBI Sequence Read Archive (SRA) is the primary sequence repository for NIH-funded high throughput sequencing activities, and the current 4.3 petabases of sequence of which 54% is open access without restriction. Computational access to SRA data, both public and controlled, is provided via the *sra-toolkit*, a software product developed by NCBI. It is freely available and runs in linux, mac and windows operating systems. Popular functions within the SRA toollkit include: *fastq-dump* to convert SRA data into FASTQ format; *prefetch* to transfer an SRA record to local disk in native format; *sam-dump* to write out SRA data in SAM format, and *sra-pileup* to generate pileup statistics for aligned SRA data, i.e. BAM submissions.

NCBI has further simplified access to SRA data by developing software extensions within the Broad Institute’s GATK genomic sequence analysis package. This presentation will demonstrate the installation and execution steps for [aligning reads and] calling variants in *Glycene max* without the burden of pre-fetching accession sequences or genome reference.

W609: NCBI Genome Resources

The NCBI Transcriptome Shotgun Assembly (TSA) Database

Susan Schafer, NIH/NLM/NCBI, Bethesda, MD

Scientists determine transcriptomes using a variety of computational techniques to assemble sequences from transcribed RNAs into longer transcripts that may represent a complete expressed gene. By analyzing an organism's transcriptome, researchers can determine expression levels of genes in different cells and tissue types to better understand health and disease.

The Transcriptome Shotgun Assembly (TSA) division of GenBank was developed specifically for high-throughput shotgun assemblies of RNASeq data. The database serves as an archive of computationally assembled transcripts from primary sequence reads deposited in the Sequence Read Archive (SRA). Sequences are deposited to TSA through the NCBI Submission Portal. Submitting rich contextual metadata along with the transcriptome data is important for providing users with a complete understanding of the source of the biological data. This information is captured in two databases designed specifically to house contextual metadata, BioProject for project-specific information and BioSample for sample-specific attributes. All data associated with a transcriptome project is linked via the BioProject ID. The data can be retrieved by both Entrez and the Shotgun Assembly Sequences browser. TSA is a resource for scientists to retrieve and analyze transcriptomic data to make important scientific discoveries.

Also, the NCBI's eukaryotic genome annotation pipeline, uses transcript sequences including TSAs for genome mapping and to compensate for gaps in genome assemblies. The pipeline automatically constructs model RefSeq transcripts and proteins (XM_ and XP_ accession prefixes) that combine genome and transcript sequence, resulting in a more accurate and complete model than is possible from using only the genome sequence.

W610: NCBI Genome Resources

Using GFF for Genome Annotation Submissions to GenBank

Terence D. Murphy, Karen Clark, Frank Ludwig, Vincent Calhoun, Sergiy Gotvyanskyy and Colleen Bollin, National Center for Biotechnology Information, NLM, NIH, Bethesda, MD

GenBank submission improvements in the last few years have made it easier for users to deposit whole genome sequence data in the public INSDC archives, which provides data standardization and greatly expands access to these important datasets. Whole genome sequence data can be more useful when it incorporates at least basic annotation information about the location of genes and proteins. The resulting annotations are available in many NCBI resources, including the nuccore, protein and BLAST NR databases, and are provided on NCBI's redesigned genomes FTP site in standardized formats, making it advantageous for submitters to provide valid annotation along with their genome sequencing data. Traditionally, annotation submission to GenBank has used a five-column feature table that is not a typical output of most annotation software. NCBI is developing tools to read annotation from GFF3 or GTF-formatted files, convert it to NCBI's standard formats, and validate the data before submission to GenBank. This talk will discuss basic requirements for using GFF3 or GTF to submit annotation, including an overview of required and optional attributes, and describe markup that will be automatically added if not available in the original submission files.

W611: NCBI Genome Resources

Recent Updates in the Eukaryotic Genome Annotation Pipeline

Francoise Thibaud-Nissen, National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD and the Eukaryotic Genome Annotation Team, NCBI/NLM/NIH, Bethesda, MD

The NCBI Eukaryotic Genome Annotation Pipeline (www.ncbi.nlm.nih.gov/genome/annotation_euk/) has been used to annotate over 260 organisms, ranging from plants to insects and mammals. The pipeline provides content for various NCBI resources including RefSeq sequence databases, Gene, BLAST databases and the Map Viewer genome browser.

The pipeline uses a modular framework for the automated execution of all annotation tasks from the fetching of raw and curated data from public repositories to the submission of the RefSeq-accessioned annotation products to public databases. The quality of the annotation is highly dependent on the availability of evidence for the species or closely related species. Alignments of RNA-Seq, traditional transcripts, ESTs, transcript assemblies and proteins by Splign and ProSplign all contribute to the prediction of gene models by Gnomon, an alignment- and HMM-based gene prediction program developed at NCBI. High-quality annotation is achieved also by weighting RefSeq curated evidence more heavily than non-curated evidence, and by producing models that compensate for assembly issues. The final products of the pipeline include the annotated genomic sequences, the genes, and the transcript and protein products named based on orthology to model organisms or Blast hits to SwissProt/UniProtKb.

We aim to re-annotate every organism we maintain every two years, so that the annotation incorporates recent evidence deposited in public databases, and benefits from improvements in the software. We will present a new type of report, to be published which each re-annotation, which categorizes the changes for each gene between two annotation releases.

See all NCBI-annotated eukaryotes at: http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all

W612: NCBI Genome Resources

Gene: Your Portal into NCBI's Eukaryotic Genome Annotations

Terence D. Murphy, National Center for Biotechnology Information, NLM, NIH, Bethesda, MD

NCBI's Gene resource (http://www.ncbi.nlm.nih.gov/gene) integrates data from multiple sources to provide a plethora of gene-specific information to aid in research and analyses. Genome annotations are taken from NCBI's RefSeq database, making it a convenient way to access annotations for over 260 eukaryotes annotated with NCBI's eukaryotic genome annotation pipeline, as well as annotation for hundreds of other eukaryotes propagated from GenBank. The content (citations, nomenclature, genomic location, gene products and their attributes, phenotypes, sequences, interactions, variation details, maps, expression, homologs, protein domains and external databases) is available via interactive browsing through NCBI's Entrez system, via NCBI's Entrez programming utilities (E-Utilities and Entrez Direct) and for bulk transfer by FTP. This talk will focus on key features of Gene, including the graphical viewer for the genome annotation which includes access to a variety of supplemental tracks including RNA-seq expression data, and is easily configured through a new, customizable Track Sets feature. Genome annotations available from Ensembl or GenBank for the same assembly are also viewable in the graphical display, and identifiers for similar annotations are provided to help integrate with other datasets. Gene also provides access to orthology information computed for over 150 vertebrates with crosslinks to corresponding human and mouse records, and this information is used to provide informative nomenclature. Data is also imported from model organism databases, the Gene Ontology Consortium, KEGG, REACTOME, and many other sources to provide a powerful portal to a wide variety of data.

W613: NCBI Genome Resources

Genomic Data Access Through BLAST

Tao Tao, NIH/NLM/NCBI, Bethesda, MD

The BLAST programs from NCBI are popular tools for sequence analysis. The BLAST web service from NCBI combines these tools with a large collection of publicly available sequences as BLAST databases to meet different research goals. NCBI also provides standalone BLAST binaries that can search custom local databases or work as clients to search databases at the NCBI. In this talk, I will summarize the types of nucleotide databases available, their organization at NCBI, and how you can select the appropriate ones for your research using the BLAST web interface (http://blast.ncbi.nlm.nih.gov). For users with high throughput needs, I will present alternatives to the BLAST web interface for searching genomic data, including using the standalone BLAST+ package, its client function, and the vdb_blastn/vdb_tblastn from NCBI's sratoolkit for next-gen reads. I will also demonstrate how to search and retrieve BLAST database information using the Entrez Utilities framework to choose the appropriate databases for these alternative BLAST avenues.

W614: New Approaches for Developing Disease Resistance in Cereals

Barley *rpg4/Rpg5* Integrated Decoy Resistance to *Ug99*; Towards Effector Identification

Robert S. Brueggeman, Roshan Sharma Poudel, Shyam Solanki and Jonathan Richards, Department of Plant Pathology, North Dakota State University, Fargo, ND

The *rpg4*/*Rpg5* locus, harbors two NLR R-genes, *Rpg5* and *HvRga1,* that are required together for resistance against wheat stem rust *Puccinia graminis* f. sp. *tritici* (*Pgt*) races including *TTKSK (Ug99).* Rpg5 is a predicted NLR resistance-proteins with an additional C-terminal kinase (STPK) domain. The transcription factor, *HvVOZ1*, was identified by yeast-two-hybrid of a library constructed from RNA of the *rpg4*/*Rpg5*+ line Q21861 48 hours post inoculation utilizing the Rpg5-STPK as bait. We hypothesize that the Rpg5-STPK acts as an integrated decoy that HvVOZ1 binds to negatively regulate defense activation or binds after activation as part of a signaling complex. The second NLR protein HvRga1 may guard the VOZ1-Rpg5 interaction or the Rpg5-STPK domain from *Pgt rpg4/Rpg5*-Avr (*r45-Avr*) effector manipulation. HvRga1 detects the manipulation by *r45-Avr* triggering the defense responses. The *r45-Avr* needs to be identified to investigate these mechanisms and test our hypothesis, thus a panel of 37 rust isolates with differential reactions on *rpg4/Rpg5* were genotyped using restriction site associated DNA-genotyping-by-sequencing (RAD-GBS) producing 4,919 informative SNPs. The genotyping information was used to select 24 diverse isolates (16 *avrRpg4/rpg5*- and *8 Avrrpg4/Rpg5*+) that are being used to conduct RNAseq during the infection process 5 days post inoculation. The RNAseq data will be used to identify *Puccinia graminis* SNPs within genes expressed during the infection process that will be added to previous RAD-GBS SNPs to conduct association mapping. This robust genotyping and the solid phenotyping on these diverse differential isolates should provide enough power to identify candidate *r45-Avr* genes.

W615: New Approaches for Developing Disease Resistance in Cereals

New Sources and Strategies for Virus Resistance in Wheat

John Fellers, USDA ARS, Manhattan, KS

Each year viruses cause significant yield loses. Wheat is no exception. Unfortunately, genetic sources of resistance are limited in availability, thus cultural control methods must be relied upon to reduce the incidence of virus infection. Wheat viruses are usually found in complexes including, but not limited to, *Wheat streak mosaic virus* (WSMV), *Triticum mosaic virus* (TriMV), *Barley yellow dwarf virus* (BYDV), and *Wheat mosaic virus*. We have screened the CIMMYT synthetic wheat collection and have identified three lines that have temperature sensitive resistance to WSMV. Two of these lines are also resistant to TriMV. We have also used two biotechnological approaches to develop resistance. Using RNAi, lines were developed expressing fragments of the WSMV CP and TriMV CP. After 6 generations, and two backcrosses into adapted varieties, lines are resistant to WSMV and TriMV. A very promising approach was taken to silence wheat eukaryotic initiation factors (eIF) 4E-2 and 4G, which are used by the viruses to complete their life cycle . Wheat lines were developed that are resistant to WSMV, TriMV, mixed infections of both, and BYDV.

W616: New Approaches for Developing Disease Resistance in Cereals

Genomic Prediction for Rusts and Leaf Spotting Diseases in Wheat – a Comparison of Models

Philomin Juliana, Cornell University, Ithaca, NY

The unceasing plant-pathogen arms race and the ephemeral nature of resistance genes has been a major challenge to breeders. One promising approach that could accelerate the gain from selection for disease resistance is 'genomic prediction' which utilizes dense genome-wide markers to estimate the breeding values for quantitative traits. While several genomic prediction models have been proposed, our objective was to compare them with a multiple linear regression(MLR) model that uses only quantitative trait loci(QTL) linked markers and also with a pedigree model. For this, we used CIMMYT's 45th International Bread Wheat Screening Nursery (267 entries) that was genotyped using Genotyping By Sequencing markers and phenotyped for different diseases in replicated trials at El Batan(Mexico), Njoro(Kenya) and Toluca(Mexico). Our results show that the mean prediction accuracies were the highest for seedling resistance to tan spot (0.72±0.08) and to leaf rust (0.64±0.11), followed by stem rust (0.58±0.06), Stagonospora nodorum blotch (0.55±0.07), stripe rust (0.48±0.14), tan spot (0.45±0.08), Septoria tritici blotch (0.44±0.1) and leaf rust (0.41±0.08). Among the models, MLR gave the lowest mean prediction accuracies (0.34±0.13), while Reproducing Kernel Hilbert Spaces(RKHS) with markers and pedigree gave the highest (0.56±0.13). The Genomic-BLUP (0.53±0.12); Genomic-BLUP mixed model (0.52±0.14); Bayesian Ridge-Regression (0.53±0.12); BayesA (0.53±0.12); BayesB (0.52±0.12); BayesCπ (0.53±0.12); Bayesian LASSO (0.52±0.12); RKHS markers (0.53±0.12) and RKHS pedigree (0.52±0.13) yielded similar accuracies. Overall, our results indicate that (i)prediction accuracies depend on the genetic architecture of resistance (ii)using genome-wide markers is advantageous than using only QTL-linked markers (iii)using both pedigree and markers results in slightly higher accuracies.

W617: New Approaches for Developing Disease Resistance in Cereals

Speed Breeding for Multiple Disease Resistance

Lee Hickey, QAAFI, The University of Queensland, BRISBANE, Australia

A new method for rapid generation advance, called 'speed breeding', has considerable advantages over DH technology for spring wheat and barley because it provides increased recombination during line development and enables selection in early generations for disease traits. The system has been refined over the past 8 years at The University of Queensland, utilizing controlled temperature regimes and 24-hour light to accelerate plant growth and development. The low-cost management system enables up to 6 generations of wheat and 5 generations of barley annually.

We have developed methods adapted for use in the speed breeding system, which permit year-round high-throughput screening for multiple disease resistance in wheat, including; biotrophic, necrotrophic and fusarium pathogens. In this presentation, we describe the protocols, explain how phenotypes relate to field-based measures and highlight how the system can also handle diverse germplasm.

We have applied similar techniques in barley. Within a two-year period, we combined speed breeding and a novel phenotypic screening protocol to transfer resistance to four foliar diseases into European cultivar Scarlett, which is widely grown throughout Argentina. Assessment of the introgression lines in disease nurseries and preliminary yield trials revealed the population was enriched with multiple disease resistance and some lines out-yielded Scarlett by 50%. Further, micro-malting results indicate elite lines display a quality profile similar to Scarlett.

This approach presents new opportunities for cereal breeders to rapidly introgress multiple disease resistance into genetic backgrounds preferred by industry, enrich breeding populations with disease resistance genes or efficiently combine target genes in top crosses.

W618: New Approaches for Developing Disease Resistance in Cereals

Genetical Microscopy: Linking the Genic Basis of Maize Disease Resistance to Effects on Pathogenesis

Randall J. Wisser, University of Delaware, Newark, DE

Unraveling the basis of complex traits and capitalizing on the knowledge gleaned from this for population improvement continues to challenge a broad community of researchers. Quantitative disease resistance is one such trait for which the level of complexity is even greater because it involves microscopic interactions between organisms. The DR Maize (Disease Resistance of Maize) project is working toward the discovery and validation of genomic variants for resistance to quantitative defense against fungal foliar pathogens. This is being pursued using a combination of resequencing, GWAS, breakpoint analysis, mutant screens, transcriptomics, and transformational validation. Simultaneously, pathogenesis is being characterized by 3D microscopy of infected host isogenic lines that contrast for natural and induced allele effects, thereby elucidating how quantitative resistance genes function to limit disease development. An experimentally validated imaging platform and computer vision tools have been developed for this. This talk will highlight multiple aspects of the project related to gene discovery and validation, while focusing on recent developments for imaging microscopic plant-pathogen interactions across a macroscopic area of host tissue and placing those observations into the context of disease development and gene function. Working with two fungal pathogens along with other evidence suggests the imaging platform is extensible to studying other plant/pathogen systems, but differences in pathogenesis introduce the need for some tailored approaches.

W619: Next Generation Genome Annotation and Analysis

Genome Sequences of Right Whales: Resource for Studies of Population History and Health

Michael S. Campbell[1], Brian K. Dalley[2], Guy Eroh[2], Andrea Chirife[3], Matias Di Martino[3], Carina Marón[4,5], Mariano Sironi[5], Luciano O. Valenzuela[4,5], Sancy A. Leachman[2], Bradley N. White[6], Marcela Uhart[3], Victoria J. Rowntree[3,4], Mark Yandell[7] and Jon Seger[4], (1)Cold Spring Harbor Laboritory, Cold Spring Harbor, NY, (2)Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, (3)Southern Right Whale Health Monitoring Program, Puerto Madryn, Chubut, Argentina, (4)Department of Biology, University of Utah, Salt Lake City, UT, (5)Instituto de Conservación de Ballenas, Ing. Maschwitz, Prov. De Buenos Aires, Argentina, (6)Department of Biology, Trent University, Peterborough, ON, Canada, (7)Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT

The world's three species of right whales (*Eubalaena* spp.) were hunted nearly to extinction. Southern right whales (*E. australis*) are recovering well, with a population of roughly 12,000 individuals, but North Atlantic (*E. glacialis*) and North Pacific (*E. japonica*) right whales are still endangered, with populations of 500 or fewer. We have sequenced four southern, three North Atlantic and one North Pacific right whales to a depth of ~40x using the Illumina TruSeq Nano DNA sample prep. For one of the southern whales, we generated 3-, 5- and 7-kb mate-pair libraries and sequenced them to a depth of ~60x. This individual was also sequenced using the Illumina Moleculo protocol, producing ~5kb (sd=2.5kb) virtual reads to a depth of ~3x. We assembled the Illumina reads with the ALLPATHS assembler and used PBjelly to fill gaps in the assembly using the Moleculo reads. The resulting assembly has a 379kb scaffold N50. CEGMA analysis of the assembly reported that 237 (95%) of the 248 ultra-conserved core eukaryotic genes are full length, with an additional 10 (4%) present but less than 70% complete, indicating the assembly includes ~99% of the genome's protein coding genes, with ~95% complete. We have annotated the assembly with the MAKER genome annotation pipeline, aligned the read data from all individuals to this reference assembly, and called variants. These data have been analyzed to estimate the nuclear mutation rate for right whales and to address questions about the historic sizes and migration histories of the three populations.

W620: Next Generation Genome Annotation and Analysis

BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs

Felipe A. Simão[1], Robert M. Waterhouse[2,3], Panagiotis Ioannidis[4], Evgenia V. Kriventseva[4] and Evgeny M. Zdobnov[4], (1)Swiss Institute of Bioinformatics, Geneva, Switzerland, (2)University of Geneva Medical School & Swiss Institute of Bioinformatics, Geneva, Switzerland, (3)The Broad Institute of MIT and Harvard, Cambridge, MA, (4)Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland

With ever-lowering sequencing costs, genome sequencing projects have been initiated for a wide range of organisms, however the vast majority of genomes currently exist in the form of draft assemblies. Considering that an important driving force behind genome projects is the acquisition of a complete catalog of genes, an important task lies in assessing the integrity and completeness of the assembled genome. Although some indications of assembly quality may be gleaned from statistical measures, a key measure of assembly quality lies in its completeness in terms of the expected gene content. The identification of genes from many diverse species that are evolving under single-copy control defines such an evolutionarily-informed expectation regarding gene content. Selected from the major species clades of the OrthoDB catalog of orthologs, the Benchmarking Universal Single-Copy Orthologs (BUSCOs) defines sets of genes expected to be present in any newly-sequenced genome from the appropriate species clade. The BUSCO assessment tool, available from http://busco.ezlab.org, implements a computational pipeline to identify and classify matches from genome assemblies, annotated gene sets, or transcriptomes. This approach provides a novel and complementary metric to assess genome assembly quality, integrity and completeness that can be easily applied to draft genome assemblies even in the absence of any gene annotations. This facilitates informative comparisons and allows the quantification of iterative improvements to assemblies or annotations. BUSCO assessments therefore offer intuitive metrics, based on evolutionarily-informed expectations of gene content from hundreds of species, to gauge completeness of rapidly-accumulating genomic data.

W621: Next Generation Genome Annotation and Analysis

Genomic insights into the (repeated) evolution of electric organs in fishes

Jason Gallant, Michigan State University, East Lansing, MI

Electric organs have evolved six or seven times in fishes to produce electric fields used in communication, navigation, predation, and in some cases, defense. In my laboratory, we are interested in two fundamental questions: the origins of what Charles Darwin once called this 'wondrous organ', and how it has diversified in form and function within each of these groups. My laboratory's recent research efforts have focused on the generation and analysis of a wide variety of genomic and transcriptomic datasets from major lineages of electric fishes using NGS

platforms. First, I will discuss our efforts to characterize similarities in gene-expression between lineages of electric fish. Our results indicate that independent lineages of electric fish have leveraged similar transcription factors and developmental and cellular pathways in the evolution of electric organs. Second, I will discuss our more recent efforts to examine genomic signatures of electric signal evolution in one lineage of electric fishes of the family Mormyridae, which produce species-specific electric discharges for communication. This diversity of electric discharges has been of great importance in the evolution of a rapidly evolved "species flock" of mormyrids in the genus *Paramormyrops*.

W622: Next Generation Genome Annotation and Analysis

Mining Genetic Variation in any Species with GEMINI

Aaron Quinlan, University of Utah, Salt Lake City, UT

A fundamental analysis challenge in studying the genetic basis of traits is isolating the handful of variants underlying a phenotype from the sea of benign, irrelevant variants. Distinguishing one variant from another is effectively impossible without context. Recognizing this need and to provide a more powerful and integrated approach for variant interpretation, we have developed GEMINI (GEnome MINIng; gemini.readthedocs.org). GEMINI combines genetic variants with a diverse and adaptable collection of genome annotations (e.g., dbSNP, ExAC, UCSC, ClinVar, 1000 Genomes) into a unified database to facilitate variant interpretation in studies of human disease. By integrating variants with extensive genome annotations, as well as sample phenotypes and family structure, one can compose complex queries based on sample genotypes, inheritance patterns, and crucial genome annotations. Owing to these strengths, GEMINI is widely-used in studies of rare human diseases.

Here we present an optimized and much more general version of GEMINI. By providing built-in annotations for model organisms and also allowing users to specify the set of annotations that are germane to their species and study needs, GEMINI has been completely redesigned to support genetic research in any species. We will discuss the functionality that GEMINI enables and demonstrate its utility in studies of genetic traits in diverse species.

W623: Non-coding RNA

Identifying and Cataloguing LncRNAs in Human and Mouse

Jennifer Harrow, Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Many groups are generating and data-mining a wealth of Illumina RNAseq data available in the public domain to identify “tens of thousands” of novel long non-coding RNAs, however their reliability is variable. It can depend on the length and quality of input data and algorithms used. As part of the GENCODE consortium we are combining different resources to produce a reference non-coding gene catalogue in human and mouse, publicly available in UCSC and Ensembl browsers. Currently we have identified around 15 900 human loci and 8 000 mouse loci that are potential long non-coding (lnc) genes. Nomenclature and classification of these entities is usually based on proximity to other coding genes, rather than based on their function. As part of the GENCODE project, we have analysed 400 lncRNAs identified as partial transcripts through the lack of CAGE data or polyadenylation signals. We have extended these loci using RACEseq and long read protocols to investigate expression in 8 different tissues. The majority of lncRNA sequences appear to be poorly conserved on the sequence level, yet annotating both mouse and human regions in parallel reveals syntenically equivalent transcripts. We are also using capture-seq technology and PacBio sequencing to compare expression of lncRNAs in the different organisms and identify novel full-length transcripts. The coding potential of these novel loci was investigated using proteomics data from the Kuster and Pandey labs. In summary, we highlight how this mix of next generation data may double the number of genes in GENCODE, presenting new challenges in cataloguing functional lncRNAs for human and mouse.

W624: Non-coding RNA

A Genomic and Transcriptomic Analysis of Factors Driving lincRNA Diversification: Lessons from Plants

Mark A Beilstein, School of Plant Sciences, University of Arizona, Tucson, AZ

Transcriptomic analyses from across eukaryotes have led to the conclusion that most of the genome is transcribed at some point in the developmental trajectory of an organism. One class of these transcripts is termed long noncoding RNAs (lncRNAs). Recently, attention has focused on understanding the evolutionary dynamics of lncRNAs, particularly their conservation within genomes. Here we take an explicitly phylogenetic approach to uncover factors influencing lncRNA emergence and persistence, both of which can affect the diversification of lncRNA populations among species. For the first time in any group of organisms, we infer phylogeny for conserved lncRNAs and use gene tree / species tree reconciliation to examine their rates of duplication, tempo of evolution, and origins. Our analyses center on >12,000 intergenic lncRNAs (lincRNAs) in the plant family Brassicaceae, a model system for evolutionary inference. We find that Brassicaceae lincRNA populations are diversifying at a rate 4-5x faster than mammalian lincRNAs over similar evolutionary timeframes. Emergence of lincRNA loci appears to be linked to local duplication events, but surprisingly, not whole genome duplication events (WGD) or the activity of transposable elements. Interestingly, WGD events impact lincRNA diversification by accelerating loss of lincRNA loci. Compared to species-specific lincRNAs, evolutionarily conserved lincRNAs are more likely to be stress-responsive, contain miRNA binding motifs, and originate from a protein-coding locus, highlighting potential functional pathways for these lincRNAs. In sum, we provide insight into the processes influencing lincRNA diversification in eukaryotes, as well as inferring and refining lincRNA datasets across the Brassicaceae.

W625: Non-coding RNA

Using Long Read Transcriptome Sequencing for LncRNA Prediction in Non-model Organisms

Richard Kuo, Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom and David W. Burt, Roslin Institute Univ of Edinburgh, Edinburgh, United Kingdom

Identifying long non-coding RNA in non-model organisms poses a major challenge due to the lack of accurate transcript models in current annotations. These non-model organism transcriptomes are compiled mostly from short read RNAseq data. Due to transcript assembly limitations, short read data can only produce abstract transcript models. Since most lncRNA prediction tools use coding potential as a criteria, full length accurate transcript models are crucial for lncRNA detection. Thus abstract models can result in a large number of false positives. Using long read sequencing can bypass these issues and open up new strategies for lncRNA discovery.

We have developed methods for identifying putative lncRNA genes using long read sequencing. From these methods we have found that current long read technology can make considerable contributions to annotating lncRNA in non-model organisms.

W626: Non-coding RNA

Discovery of Conserved RNA Secondary Structures in Farm Animals

Stefan E. Seemann, Center for non-coding RNA in Technology and Health, Department of Veterinary Clinical and Animal Sciences, University of Copenhagen, Frederiksberg C, Denmark

The molecular structure of especially long RNA transcripts is still largely unexplored, but often directly associated with their functionality. For instance, structured RNA domains are involved in binding nucleic acids and proteins as well as RNA stabilization. Recent transcriptome-wide in-vitro structure probing experiments have revealed the complexity of the structure landscape; however, these attempts do not consider that functional important RNA domains will be under phylogenetic constraints.

We present a large collection of RNA secondary structures conserved in vertebrates (CRSs) that we computationally predicted in a genome-wide screen using human as the reference organism. CRSs overlap a range of known sequences, from Rfam to long non-coding RNAs, and with most of the CRSs in intronic or intergenic regions. Many of the RNA structures are under purifying selection, despite of that many have low sequence identity (<50%). Comparing the CRSs to causal autoimmune disease variants, we predict several SNPs to cause changes in the RNA structure. CRSs are significantly co-localized with RNA binding sites of many proteins and cis-regulatory regions. The latter comprise enhancer and promoter regions which have recently been shown to be transcribed as eRNAs and PROMPTs. A large fraction of the CRSs are located in transcribed genomic regions, and we show evidence that their expression is conserved in human and mouse by analyzing total RNAseq and qPCR. Finally, we use conservation on the level of RNA structure, rather than primary sequence, to discover structured RNAs in farm animals.

W627: Non-coding RNA

Positionally-Conserved but Sequence-Diverged: Identification of Long Non-Coding RNAs in the Brassicaceae and Cleomaceae

M Eric Schranz, Wageningen University and Research, Wageningen, Netherlands

Long non-coding RNAs (LncRNAs) have been identified as gene regulatory elements that influence the transcription of their neighbouring protein-coding genes. The discovery of LncRNAs in animals has stimulated genome-wide scans for these elements across plant genomes. Recently, 6480 LincRNAs were putatively identified in *Arabidopsis thaliana* (Brassicaceae), however there is limited information on their conservation. Using a phylogenomics approach, we assessed the positional and sequence conservation of these LncRNAs by analyzing the genomes of the basal Brassicaceae species *Aethionema arabicum* and *Tarenaya hassleriana* of the sister-family Cleomaceae. Furthermore, we generated transcriptomes for another three *Aethionema* species and one other Cleomaceae species to validate their transcriptional activity. We show that a subset of LncRNAs are highly diverged at the nucleotide level, but conserved by position (syntenic). Positionally conserved LncRNAs that are expressed neighbour important developmental and physiological genes. Interestingly, >65 % of the positionally conserved LncRNAs are located within 2.5 Mb of telomeres in *Arabidopsis thaliana* chromosomes. These results highlight the importance of analysing not only sequence conservation, but also positional conservation of non-coding genetic elements in plants including LncRNAs.

W628: Non-coding RNA

Assessing the Ability of RNA-Seq to Quantify Non-Coding RNA Accurately

Mick Watson, The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom

Transcriptomics is an important approach that has helped researchers understand the molecular basis of disease in a range of species. Whilst for many years microarrays were the tool of choice, RNA-Seq has now emerged as the standard method for analysing the transcriptome, contributing to thousands of publications in the biomedical literature. However, the vast majority of RNA-Seq analysis has focused on protein-coding genes, and we do not understand how well different software tools perform in the quantification of non-coding RNAs. Farm animal genomes lag behind the human genome in terms of non-coding RNA annotation, so we focused on the human genome for this study. We simulated RNA-Seq reads from human non-coding RNA and ran several RNA-Seq quantification pipelines to find out which non-coding genes can be accurately quantified by RNA-Seq

W629: Non-Seed Plants

The Porphyra Genome Project

Simon Prochnik, DOE Joint Genome Institute, Walnut, Creek, CA

W630: Non-Seed Plants

The Anthocereos Genome Project

Peter Szovenyi, University of Zurich, Zurich, Switzerland

The monophyletic group of hornworts is believed to represent the immediate sister group of all vascular land plants. However, this traditional view is still debated and cannot be satisfactorily resolved owing to the lack of detailed knowledge on the general biology and genomic features of hornworts. Until now, advancement in this field was primarily hindered by the lack of a hornwort model species appropriate for experimental work. Here we describe the establishment of a tractable model species for hornworts, *Anthoceros agrestis*, and its genomic features. We provide detailed protocols to grow *A. agrestis* throughout its life cycle in axenic culture. Furthermore, we show that *A. agrestis* has a remarkably small genome, with few recent paralogs, which makes it approriate for genetic analysis. Finally, we report on the genomic features of the chloroplast, mitochondrion and nuclear genomes and compare those with algal and vascular land plant genomes. We will also summarize our achievements and provide a list of issues that need to be resolved in the future.

W631: Non-Seed Plants

Reconstructing the Ancestral Gene Set of Bryophytes from Comparative Transcriptome Data

Matthew Johnson[1], Kristian K. Ullrich[2], Stefan A. Rensing[3] and Norman Wickett[1], (1)Chicago Botanic Garden, Glencoe, IL, (2)Department of Cell Biology - University of Marburg, Marburg, Germany, (3)Faculty of Biology, University of Marburg, Marburg, Germany

Transcriptome data generated for the 1KP Project has greatly increased the genomic resources for non-model organisms, especially bryophytes. Protein orthology in the 1KP Project does not account for gene families that are restricted to specific clades if the ortholog in a reference genome is single-copy. We describe a process for identifying valid protein families that are unique to bryophytes but absent in the sequenced genomes of the moss *Physcomitrella* and the liverwort *Jungermannia*. Reclustering of proteins not previously sorted into orthogroups reveals thousands of gene families that may be unique to the bryophytes among land plants. By identifying orthologous proteins across many species, we can rule out contamination and distinguish ancient gene families that have been lost in model species from novel gene families restricted to specific clades. Using these orthogroups we explore patterns of gene family expansion and loss of ancestral gene families within bryophytes with a focus on identifying gene families that may have originated through horizontal transfer from bacteria or fungi.

W632: Non-Seed Plants

Explorations into the C-Fern Genome

Blaine Marchant[1], Brad Barbazuk[2], Emily Sessa[2], Matias Kirst[2], Pamela S. Soltis[2] and Douglas E. Soltis[2], (1)University of Florida, Gainsville, FL, (2)University of Florida, Gainesville, FL

In a little more than a decade, our understanding of plant genomes has increased immensely. Sequenced genomes are now available for every major green plant clade, from chlorophytic and streptophytic algae to a vast array of flowering plants - that is, every clade except the monilophytes (ferns). Ferns are the most biodiverse clade of land plants after the angiosperms and sister group to the economically significant seed plants. Notorious for large genomes (~12 Gb) and numerous chromosomes (~3x more than the average angiosperm), fern genomes are largely unexplored despite their evolutionarily significant position as a reference group for analyzing ancestral versus derived genomic and genetic characters in the seed plants. Using transcriptomic, genomic, and chromosome fluorescent in situ hybridization techniques, we are delving into the genome of *Ceratopteris richardii*, a fast-growing tropical aquatic homosporous fern with a haploid genome size of 11.26 Gb and chromosome number of $n = 39$. We will present our results regarding gene specificity in the gametophytic and sporophytic life stages, gene density, and the role of polyploidy, retrotransposons, and small-scale duplications in the evolutionary genomics of this species. The outcome of this research will provide novel perspectives on the genomic dynamics and characteristics of this major clade, while also providing crucial resources for broader comparative genomic, phylogenetic, and developmental studies between ferns and seed plants, yielding new insights into the evolution of euphyllophytes as a whole.

W633: Non-Seed Plants

The Azolla Genome Project

Andrea Braeutigam[1], Nils Koppers[2], Mathew Simenc[3], Fay-Wei Li[4], Laura Dijkhuizen[5], Paul Brouwer[5], Shifeng Cheng[6], Xin Liu[6], Bo Song[6], Gane Ka-Shu Wong[7], Andreas P. M Weber[8], Kathleen M. Pryer[9], Joshua P. Der[3] and Henriette Schluepmann[5], (1)Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany, (2)Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany, (3)California State University, Fullerton, Fullerton, CA, (4)University of California Berkeley, Berkeley, CA, (5)Utrecht University, Utrecht, Netherlands, (6)Beijing Genomics Institute-Shenzhen, Shenzhen, China, (7)Department of Biological Sciences, University of Alberta, Edmonton, AB, AB, Canada, (8)Plant Biochemistry (CEPLAS) Heinrich Heine University Düsseldorf, Duesseldorf, Germany, (9)Duke University, Durham, NC

Azolla filiculoides (Salviniales) is a floating aquatic fern capable of fixing nitrogen from the air because of its obligate symbiotic association with the cyanobacterium *Nostoc azollae*, allowing it to achieve high growth rates even in the absence of ammonium or nitrates in the medium. The *Azolla* genome fills a critical phylogenetic gap among the five major embryophyte lineages; reference genomes are already available for bryophytes (*Physcomitrella*), lycophytes (*Selaginella*), gymnosperms (*Picea* and *Pinus*), and flowering plants (several representatives). In addition, the *Azolla* genome will inform us about the plant-cyanobacterial symbiotic partnership and the accompanying bacteria in the leaf pocket. *Azolla* was sequenced using a PacBio-only strategy supplemented with genome sequencing and transcriptome sequencing by Illumina and IonTorrent. Reads from 43 SMRT cells (51x coverage) were error-corrected with PBcR resulting in 23.5 Gb to which 95.9% of the assembled transcripts could be mapped. Different assembly options were tested before corrected reads were reduced to 25x coverage and assembled with Celera resulting in 693Mb of the genome (estimated at 750Mb). The genome was assembled into 4737 contigs with an N50 of 0.89Mb, and an average contig length of 146kb with the longest contig at 1.65Mb. 97% of the Illumina genome reads and 95.9% of the assembled transcripts mapped to the assembly, which has been scaffolded using BioNano technology. We will report on the ongoing genome annotation, particularly gene content and duplication status, and its implications for the evolutionary dynamics that are fundamental to plant life, as well as to the plant-cyanobacterial symbiosis.

W634: Non-Seed Plants

Evo-Devo of Leaves: A Story Told by Ferns and Lycophytes

Barbara A. Ambrose, The New York Botanical Garden, Bronx, NY

Biologists have been intrigued by the origin of novelties for centuries. Leaves are novel plant organs and can serve as an excellent model for understanding the origin and evolution of novelties. Morphological, paleobotanical, and phylogenetic analyses have failed to concur on the origin of leaves or even how many times leaves have evolved in vascular plants. In addition, previous comparative expression analyses in lycophytes and seed plants have come to diametrically opposed conclusions about the conservation of a leaf developmental program. We conducted phylogenetic analyses of leaf developmental gene families (Class III HD-Zip, KNOX, CUC) across land plants with extensive new data across lycophytes and ferns. We also utilized comparative expression analyses of orthologous genes by *in situ* hybridization. We performed our expression analyses in both vegetative and reproductive organs of lycophytes and ferns. Here, we provide the first molecular genetic support indicating a conservation in a leaf developmental program between ferns and seed plants. In addition, our results show Class III HD-Zip expression in sporangia (reproductive organs) across lycophytes and ferns. This provides the first molecular genetic evidence that a sporangia

developmental network was co-opted independently for the evolution of all leaves. Furthermore this shifts the paradigm of lycophyte leaf evolution from the widely cited enation theory to the sterilization of sporangia theory.

W635: NRSP-8 Animal Genome

Approaches Taken, Progress Made, and Enhanced Utility of Long Read-based Goat, Swine, Cattle and Sheep Reference Genomes

Timothy P.L. Smith[1], Sergey Koren[2], Adam Phillippy[2], Derek M. Bickhart[3], Benjamin D. Rosen[3], Kim C. Worley[4], Steven G. Schroeder[3], Brian L. Sayre[5], Ivan Liachko[6], Shawn T. Sullivan[7], Joshua N. Burton[6], Alex R. Hastie[8], Tad S. Sonstegard[3], Christy M. Kelley[9], Alan L. Archibald[10], Mick Watson[11], Richard Green[12], Jason Chin[13], Dan Nonneman[1], Gary A. Rohrer[1], Christopher K. Tuggle[14], Haibo Liu[15], Daniel C. Ciobanu[16], Juan F. Medrano[17], Aleksey Zimin[18], Sebastian J. Schultheiss[19], Darren E. Hagen[20], Christine G. Elsik[20], Brian Dalrymple[21], James W. Kijas[22], Noelle Cockett[23] and Michael P. Heaton[24], (1)USDA, ARS, USMARC, Clay Center, NE, (2)National Human Genome Research Institute, Bethesda, MD, (3)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (4)Baylor College of Medicine, Houston, TX, (5)Virginia State University, Petersburg, VA, (6)University of Washington - Department of Genome Sciences, Seattle, WA, (7)Phase Genomics, Seattle, WA, (8)BioNano Genomics, San Diego, CA, (9)USDA ARS US Meat Animal Research Center, Clay Center, NE, (10)The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom, (11)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (12)University of California, Santa Cruz, Santa Cruz, CA, (13)Pacific Biosciences, Menlo Park, CA, (14)Iowa State University, Ames, IA, (15)Bioinformatics and Computational Biology Program, Department of Animal Science, Iowa State University, Ames, IA, (16)University of Nebraska - Lincoln, Lincoln, NE, (17)University of California-Davis, Davis, CA, (18)University of Maryland, College Park, MD, (19)Computomics GmbH & Co. KG, Tuebingen, Germany, (20)delete, Columbia, MO, (21)CSIRO Agriculture, St. Lucia, QLD, Australia, (22)CSIRO Animal, Health and Food Science, St Lucia, Australia, (23)Utah State University, Logan, UT, (24)USDA, ARS, U.S. Meat Animal Research Center (USMARC), Clay Center, NE

Deficiencies in the current livestock reference genomes continue to hamper genomic analysis in these species, despite the extensive efforts to create, re-create, update and curate them. On the other hand, a change of reference genomes incurs substantial cost due to the need to reanalyze data and account for the new information and reference position. In livestock genomics research, which has limited resources compared to the human or biomedical model species communities, it is critical that updates to be termed a “new reference” be of sufficiently increased quality, stability, and utility that the many annotation analyses are worth re-doing. Recent advances in long-read sequencing combined with new technologies for scaffolding the resulting contigs, have made it possible to make a step change in the quality of genome assemblies for a very small fraction of the price required to create the originals. Efforts to improve the goat, swine, cattle and sheep genomes through long read-based *de novo* assemblies scaffolded with a variety of approaches are in various stages of production. We believe that these new resources represent sufficient improvement to be worth the effort to transport and enhance annotation on these new assemblies. A brief update on the status of these genomes upgrade efforts will be presented, details of which will be available in other presentations. We then follow with examples of ongoing projects making use of the preliminary outputs, to illustrate that the switch to these new references will be worthwhile.

W636: NRSP-8 Animal Genome

The Role of Mobile Genetic Elements in the Bovine Genome

Carole Charlier, University of Liège, Liège, Belgium

W637: NRSP-8 Animal Genome

Genomics of Response to Environmental Challenges of Poultry

Susan J. Lamont, Iowa State University, Ames, IA

Provision of high-quality, animal-based foods to meet the rapidly increasing demands of the world population will require increased efficiency of production. To achieve the goal of a food-secure world, the losses due to disease and environmental stressors must be reduced. One component of a diverse portfolio of strategies that can be used to address the urgent need for increased efficiency is genetic improvement of farm animals. To knowledgeably enhance the resilience of livestock and poultry populations, increased understanding is required of the diverse mechanisms and the genetic control of host response to important pathogens and to abiotic stressors such as heat.

Developing a greater knowledge base of the host genomic response can be based upon many levels of biological complexity, including isolated cells, individual tissues, whole animals, inbred lines, experimental populations and commercial lines. Studying diverse types of biotic challenges, including parasites, bacteria and viruses, as well as abiotic challenges such as heat stress, helps to elucidate those genomic responses that are relatively specific to a perturbation versus those that are conserved across a wide range of different challenges. Additionally, using analyses that capture deep phenotypes, gene expression and genetic structural variation each give different insights into the relationship of genetic variation and important biological traits. Examples will be presented of studies from the author’s and collaborators’ labs on genomics of response to pathogens and to heat stress in poultry.

W638: NRSP-8 Animal Genome

Comparative Studies of Mammalian Sex Chromosomes: From Cytogenetics to NGS

Terje Raudsepp, Texas A&M University, College Station, TX

It is a common knowledge that sex chromosome mutations are better tolerated and more viable compared to changes in autosomes. This is explained by relatively low gene density in both the X and the Y chromosome and by random X chromosome inactivation in mammalian females buffering the effect of X-aneuploidies. However, it is not well understood why apparently similar sex chromosome abnormalities, such as X-monosomy or certain Y chromosome rearrangements, result in different phenotypic effects in different species. It is thought that this is due to species differences in the organization of the Y chromosome, differences in the set of genes escaping X-inactivation, and the presence of

species/lineage specific sex-linked genes with functions in development and reproduction. Current knowledge about the species differences in sex chromosome organization and function is limited, this despite the availability of reference genome assemblies for most domestic species. It appears that sequence assembly of the X chromosome in most species is rather patchy containing multiple gaps and possible misassemblies, being the poorest in the pseudoautosomal region and in regions containing putative lineage-specific sequences. The Y chromosome, on the other hand, is typically not included in the reference genome and is studied separately, whereas complete sequence assembly of the male-specific portion of the Y is not yet available for any domestic species. In this talk I will discuss comparative organization and function of animal sex chromosomes and related phenotypes proceeding from our research in horses.

W639: NRSP-8 Animal Genome

FAANG update

Christopher K. Tuggle, Iowa State University, Ames, IA

W640: Oats

Genotyping to integrate historical research efforts in oat

Jean-Luc Jannink, USDA-ARS / Cornell University, Ithaca, NY, Nicholas A. Tinker, Agriculture and Agri-Food Canada, Ottawa, ON, Canada, Shiaoman Chao, USDA-ARS, Fargo, ND, Clare Saied, Cornell University, Dept. of Plant Breeding and Genetics, Ithaca, NY and David Matthews, Cornell University, Ithaca, NY

The oat breeding and genetics community has benefitted from two long standing cooperative resources, the Uniform Oat Performance Nurseries (UOPN) and the Pedigrees of Oat Lines (POOL) database. Data from these resources have recently been consolidated in the oat instance of The Triticeae Toolbox (T3/oat, available at triticeaetoolbox.org/oat). In the past year, a Public Oat Genotyping Initiative (POGI) started to obtain marker information on primarily public sector oat breeding lines using genotyping by sequencing (GBS). In this talk, we will present these three resources and evaluate analyses that bring together the information from them to predict the performance of experimental oat breeding lines.

W641: Oats

Haplotag: New GBS Software Facilitates Haplotype-Based GWAS in Hexaploid Oat

Wubishet Abebe Bekele, Agriculture and Agri-Food Canada, Ottawa Research and Development Centre, Ottawa, ON, Canada and Nicholas A. Tinker, Agriculture and Agri-Food Canada, Ottawa, ON, Canada

Genotyping-by-sequencing (GBS) is becoming an integral part of many plant genotyping projects. Haplotag is a population based GBS software that determines the proportion of sequence-tag-level haplotypes and selects the best model that fits the expectations of the population under investigation. In sequence tags that contain multiple SNPs, Haplotag can alternately report data either as haplotype calls or as individual SNP calls. Since multiple SNPs are often located on the same tag, and since only a subset of possible haplotypes are usually present in a population, the haplotype-based data represents the same genetic information using fewer degrees of freedom in a global association analysis. It is also possible that haplotype-based GWAS analysis may identify cryptic associations where an adapted allele is uniquely associated with an intermediate haplotype. The aim of this study was to compare GWAS using haplotype markers vs SNP markers. Results of this comparison suggest that haplotype-based GWAS analysis is more powerful than SNP-based GWAS, especially when error control is based on the stringent Bonferroni correction. An example of haplotype-based GWAS in flowering time and its hypothetical use in applied oat breeding will be illustrated.

W642: Oats

From Diploid Avena Sequence to Hexaploid Genomics and Allele Discovery: Triple the Fun?

Tim Langdon, IBERS, Aberystwyth University, Aberystwyth, United Kingdom

An Avena reference genome was constructed, based on approximately 40-fold sequencing coverage of the wild diploid *Avena atlantica* with map information provided by survey sequencing of selected progeny of a population of *A. atlantic*a x *A. strigosa*. Homology evidence and gene prediction were used to annotate the assembly with gene models. Transcriptome data from eleven *A. atlantica* tissues was used to further catalogue gene content and expression. This provides an overview of a typical wild Avena genome and gives some indication of the specific pathways and adaptations that distinguish oat from other cereals. We are currently using this information to guide assembly of genomic information from the hexaploid oat *A. sativa*, and to identify markers and candidate genes for agronomically important traits in the oat crop. A novel chromosome sorting method has greatly improved the efficiency of our hexaploid sequencing and is being applied to the common parent in our Nested Association Mapping (NAM) population. Genotyping-by-sequence and survey sequencing of NAM progeny will allow mapping of the hexaploid assemblies, and simplify association of traits and candidate genes in this population. Developing grain transcriptomes of the NAM parents are also being used to examine variation in expression in the hexaploid. The combined sequence and genotyping resources shed light on the processes of oat domestication and will help access useful genetic variation from historic and wild material.

W643: Oats

Relating Vitamin E Pathway Gene Homeolog Expression to Tocol Accumulation in Oats

Juan J Gutierrez-Gonzalez, University of Minnesota, Saint Paul, MN

Among the cereal grains, hexaploid oats (*Avena sativa* L.) are particularly rich in vitamin E, an essential lipid-soluble vitamin that maintains membrane stability and possesses antioxidant and anti-inflammatory properties. To date, no gene sequences involved in vitamin E biosynthesis have been reported for oats. We used deep sequencing and an orthology-guided assembly to reconstruct coding sequences of genes for each step in the vitamin E synthesis pathway, including resolution of the sequences of homeologs. Three homeologs, presumed to represent each of the three oat subgenomes, were identified for the main steps of the pathway. Pairwise comparisons among homeologs revealed that, for each gene, two of the three putative subgenome-specific homeologs are very similar. We estimated divergence times between the three oat subgenomes and used this information to shed light on different evolutionary scenarios that may have led to the emergence of hexaploid oat. Homeolog-specific gene expression was quantified during oat seed development and compared with vitamin E accumulation. Homeolog expression largely appears

to be similar for most of genes; however, for some genes homoeolog-specific transcriptional bias was observed. Our findings expand our understanding of oat genome evolution, and will assist efforts to modify vitamin E content and composition in oats.

W644: Oats

Homeoallelic Variation in Oat Hemicellulose Biosynthesis Genes

Alexander E. Chu[1], Melissa C. Fogarty[1], Peter J. Maughan[1], Eric W. Jackson[2] and Eric N. Jellen[1], (1)Brigham Young University, Provo, UT, (2)General Mills, Kannapolis, NC

In seeds of oat (*Avena* spp., x = 7), (1,3;1,4) β-D-glucans constitute a fundamental component of the endosperm and are a primary determinant of soluble fiber content. Precursors for β-glucan biosynthesis are derived from primary starch and cellulose metabolism. ADP-glucose pyrophosphorylase (AGPase), encoded by the *AGP* gene, is responsible for the first committed step in ADP-glucose biosynthesis. UDP-glucose pyrophosphorylase (UGPase), encoded by the *UGP* gene, catalyzes the reversible conversion of glucose-1-phosphate to UDP-glucose. Products of the cellulose synthase A6 gene (*CESA6*) are implicated in cellulose deposition in the secondary cell wall during seed endosperm development. Here, we present the sequences of the A, C, and D genome alleles of *AGP*, *UGP*, and the *CESA6* genes in six cultivars of *A. sativa*, as well as their orthologs in *A. wiestii* (AA), *A. ventricosa* (CC), and *A. canariensis* (putative DD). We further present data on alternative splice sites and differential expression of homeoalleles and discuss the implications of our data for characterizing polysaccharide biosynthesis and deposition in the oat grain, molecular breeding approaches, and understanding the nature of epistatic homeoallelic interactions in complex polyploids.

W645: Oats

International Engagement through Oat Global, T3/Oat, and the Oat Newsletter

Haiyan Jia[1], Nicholas A. Tinker[2], Jean-Luc Jannink[3], Charlene P. Wight[2], Clare Saied[4], Gerard R. Lazo[5] and Gabe Gusmini[6], (1)PepsiCo, St Paul, MN, (2)Agriculture and Agri-Food Canada, Ottawa, ON, Canada, (3)USDA-ARS, Ithaca, NY, (4)Cornell University, Dept. of Plant Breeding and Genetics, Ithaca, NY, (5)USDA Agricultural Research Service, WRRC, Albany, CA, (6)PepsiCo, St. Paul, MN

Oat is an important global food crop supported by relatively small and regional research budgets. Many challenges in oat research, oat production, and oat utilization are best met through international cooperation. In 2014, Oat Global (www.oatglobal.org) was created as a strategy forum driven by the leaders of the public and private oat community at large. Oat Global operates across agencies, institutions, corporations, and national boundaries in support of pre-competitive R&D in any research field related to oats. Oat Global provides a unifying platform for open knowledge, open data, strategy, and resource sharing. Oat Global is governed by a Strategy Committee representing an open global membership. As needs are identified, Oat Global will engage with potential funding groups to co-ordinate or motivate project teams to execute sponsored initiatives. The first Oat Global sponsored initiative was the expansion of the T3 relational database (triticeaetoolbox.org/oat/) for oat genotype and phenotype data. In its first year, Oat Global has also engaged in the development of an oat rust initiative, a public genotyping initiative, and an oat genome sequencing project. Furthermore, Oat Global has encouraged and supported the revitalization of the Oat Newsletter (oatnews.org) a forum for rapid publication of new research as well as a community news portal for oat workers. Oat Global and these new resources will be described in a short presentation. This presentation will be followed by an open workshop discussion of gaps and opportunities in oat research.

W646: ONE HEALTH Epigenomics: From Soil to People

Glyphosate and other Contaminants of Concern in Water, Sediment, and Soil in the United States

William Battaglin, U.S. Geological Survey Colorado Water Science Center, Denver, CO, Paul Bradley, U.S. Geological Survey, Columbia, MO and Michael Meyer, U.S. Geological Survey, Lawrence, KS

Pesticides, hormones, pharmaceuticals, and other contaminants of emerging concern (CEC) are released into the environment by agriculture, industry, and other human activities. The corresponding human and environmental risk is cumulative and can include individual and community-level impacts caused by toxicity, endocrine disruption, anti-microbial resistance, behavior alteration, and other sub-lethal effects. CEC are observed both in highly modified landscapes and in remote or pristine landscapes. Herbicides that utilize glyphosate as an active ingredient are of particular concern in the United States (US) largely due to their heavy use (~100 million kilograms per year) on soybean and corn crops that are genetically modified to tolerate it. However, glyphosate is seldom included in environmental monitoring programs, due in part to technical difficulties in measuring it at concentrations relevant to environmental studies. The U.S. Geological Survey (USGS) has conducted the largest and most comprehensive assessment of the environmental occurrence of glyphosate and its primary degradation product AMPA (aminomethylphosphonic acid) in the US, summarizing the results of 3,732 environmental samples collected from 1,341 sites in 38 States between 2001 and 2010. Results indicate that glyphosate and AMPA are mobile and occur widely in the environment. Glyphosate was detected in over 50% of samples from streams and rivers, precipitation, ditches and drains, and soil and sediment; but in less than 40% of samples from soil pore-water, groundwater, and wetlands.

Other CEC such as hormones and pharmaceuticals are commonly detected in surface water and bed sediment in urban and suburban areas due to the proximity and/or intensity of their use or their association with wastewater discharges and urban runoff. However, water and sediment from urbanized or agriculturally dominated landscapes are not the only place were CEC are found. In 2012-2013, we sampled water and sediment from 20 sites in Rocky Mountain National Park. Water samples were analyzed for 19 hormonally active compounds, 109 pharmaceutical related compounds, and 69 wastewater related compounds; and sediment samples were analyzed for 19 hormonally active compounds and 57 wastewater related compounds. In all, 54 compounds were detected in one or more water samples and 45 were detected in one or more sediment samples. Some detected compounds, such as carbaryl, 17-α-estradiol, celecoxib, and oxycodone, are attributable to direct human input, whereas others, such as camphor, estrone, 3-beta-coprostanol, and p-cresol, may have local natural sources such as wildlife or pine needles.

W647: ONE HEALTH Epigenomics: From Soil to People

Epigenetic Mechanisms Associated with Endocrine Disrupting Chemicals (EDCs): Their Association with Disease

Acacia Alcivar-Warren, Environmental Genomics, Inc., Southborough, MA

Information on transgenerational epigenetic inheritance (TgEI) and epigenetic mechanisms associated with EDCs will be presented, in light of globalization and climate change. EDCs are associated with DNA methylation changes, histone modifications (acetylation, methylation, phosphorylation, sumoylation, ubiquitination), RNAi mechanisms (microRNAS and other non-coding RNAs). Exposure to environmental contaminants of concern (COCs) such as EDCs (PCB pesticides, heavy metals, PAHs, herbicides, antibiotics), bacterial transgenes (*Bacillus thuringiensis*) and other stressors, can cause long-term adverse health effects to environment, animals and people. In animal models, EDCs produce abnormal reproductive or metabolic phenotypes that are transgenerationally transmitted, increasing incidence of obesity, polycystic ovary syndrome, pregnancy defects, germ cell apoptosis. EDCs are associated with chronic conditions (diabetes, cardiovascular, renal diseases, congenital malformations. When EDCs introduce epigenetic changes during early development, they permanently alter the epigenome in the germline, and the changes can be transmitted to subsequent generations. If epigenetic changes are introduced during adulthood, the changes within an individual occur in somatic cells and are not permanent or transmitted to subsequent generations (Gore et al. 2015). Siklenka et al. (2015) reported paternal epigenetic transmission of a transgene, both intergenerational effects (F1) and intragenerational effects (F2, F3). Evidence predicts TgEI of current COCs in humans will severely impact the incidence of non-infectious diseases in future generations, due to long-lasting alterations in the gametes epigenome. Limited information has been published for Glyphosate, the most used herbicide worldwide, an EDC, bactericide and metal chelator. Host–microbe interactions are also capable of mediating TgEI of a stress-induced Drosophila phenotype.

W648: ONE HEALTH Epigenomics: From Soil to People

Vibrio parahaemolyticus: An Opportunistic Marine Pathogen Becomes Virulent by Acquiring a Plasmid that Expresses a Deadly Toxin

Chung-Te Lee[1,2], I-Tung Chen[1], Han-Ching Wang[3], Hao-Ching Wang[4], Lien-I Hor[2] and **Chu-Fang Lo**[1], (1)Institute of Bioinformatics and Biosignal Transduction, College of Bioscience and Biotechnology, National Cheng Kung University, Tainan, Taiwan, (2)Department of Microbiology and Immunology, College of Medicine, National Cheng Kung University, Tainan, Taiwan, (3)Institute of Biotechnology, College of Bioscience and Biotechnology, National Cheng Kung University, Tainan, Taiwan, (4)Graduate Institute of Translational Medicine, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

Outbreaks of acute hepatopancreatic necrosis disease (AHPND) have been causing losses to the shrimp farming industry globally since 2009. The causative agent of AHPND is a specific strain of *Vibrio parahaemolyticus* that has become highly virulent by acquiring a unique AHPND-associated plasmid (pVA1). This virulence plasmid is stably inherited *via* a postsegregational killing system and it is disseminated by conjugative transfer. The plasmid encodes a binary toxin that is formed from *V. parahaemolyticus Photorhabdus* insect-related toxins $PirA^{vp}$ and $PirB^{vp}$. Neither $PirA^{vp}$ nor $PirB^{vp}$ alone are sufficient to produce the characteristic symptoms of AHPND, but the $PirAB^{vp}$ complex induces cell death, and its cytotoxicity is analogous to that of the structurally similar insecticidal pore-forming Cry toxin. Further evidence of the importance of the $PirAB^{vp}$ toxin in causing AHPND comes from *V. parahaemolyticus* strain M2-36, which was isolated from a culture pond in Vietnam after an outbreak of AHPND. Although M2-36 has acquired a pVA1-like plasmid, it fails to induce AHPND in shrimp, and we found that the reason for this was that the entire $pirAB^{vp}$ operon was missing from the M2-36 pVA1-like plasmid. The gene organization of pVA1 further suggests that $PirAB^{vp}$ may be lost or acquired by horizontal gene transfer via transposition or homologous recombination. Taken together, these results imply that any measures that successfully block the formation of the active $PirAB^{vp}$ complex and/or its interaction with $PirAB^{vp}$ receptors should greatly reduce the impact of this disease.

W649: ONE HEALTH Epigenomics: From Soil to People

Horizontal Gene Transfer (HGT) and Transgenerational Epigenetic Inheritance (TgEI): The Transposases of AHPND-causing Vibrio parahaemolyticus and *V. harveyi, Bacillus thuringiensis*, and the Pacific Whiteleg Shrimp, *Litopenaeus vannam*

Acacia Alcivar-Warren[1], Sonia Soto-Rodríguez[2], Jianhai Xiang[3], Raquel Silveira[4], Gustavo Arencibia Carballo[4], Chinnaiah Amutha[5] and Bruno Gomez-Gil[6], (1)Environmental Genomics, Inc., Southborough, MA, (2)CIAD, A.C. Unidad Mazatlán, Mazatlán, Sinaloa, Mexico, (3)Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, (4)Centro de Investigaciones Pesqueras, La Habana, Cuba, (5)Madurai Kamaraj University, Tamilnadu, India, (6)CIAD, A.C. Unidad Mazatlán, Mazatlán, Mexico

Host-microbe interactions are capable of transgenerational epigenetic inheritance (TgEI) of stress-induced phenotypes. For example, exposure to antibiotic G418 induced TgEI phenotypes, including a delay in larval development, gene induction in the gut and morphological changes (Fridmann-Sirkis et al. 2014). G418 selectively depleted commensal Acetobacter sp and the depletion explained the heritable delay, but not the inheritance of other phenotypes. TgEI mechanisms could also be associated with Glyphosate, a bactericide and EDC present in genetically modified crops (GMCs) used to produce aquafeeds. Increased commercial use of genetically engineered (GE), GMCs or GMOs has renewed calls to assess their potential ecological and health risks, new zoonosis. We review the likelihood of tolerance of insects (and non-target species like shrimp) to insect-resistant GMCs, the health effects caused by microbial transgenes (Bacillus thuringiensis, Bt); and potential impact of Glyphosate, GE, GMCs, GMOs, Bt on biodiversity, in light of globalization, climate change, HGT and TgEI. We compare conservedprotein domains of bacterial transposases of Tn3, Tn4, Tn5 transposons, Bt transposase, and transposase of new Tn6264 transposon in plasmids of Vibrio parahaemolyticus (Vp) and V. harveyi-causing Acute Hepatopancreatic Necrotic Disease (AHPND). Sequence variations of horizontally-transferred TnpA transposase (E. coli Tn1000) into L. vannamei genome will be reviewed. Transposases of Mariner, EnSpm, other transposons of L. vannamei will be presented. This review represents background information for testing the hypothesis by 'The Shrimp Epigenome (shrimpENCODE) Consortium' collaborators that AHPND/EMS is TgEI- and HGT-mediated, by both genomic and epigenetic mechanisms involving interactions of vibrio-shrimp genomes and environmental and nutritional factors.

W650: ONE HEALTH Epigenomics: From Soil to People

Laboratory Rodent Diets Contain Toxic Levels of Environmental Contaminants: Implications for Regulatory Tests with a Focus on GMOs and Pesticides

Nicolas Defarge, CRIIGEN, Caen, France
The quality of diets in rodent feeding trials is crucial. We describe the contamination with environmental pollutants of 13 laboratory rodent diets from 5 continents. Measurements were performed using accredited methodologies. All diets were contaminated with pesticides (1-6 out of 262 measured), heavy metals (2-3 out of 4, mostly lead and cadmium), PCDD/Fs (1-13 out of 17) and PCBs (5-15 out of 18). Out of 22 GMOs tested for, Roundup-tolerant GMOs were the most frequently detected, constituting up to 48% of the diet. The main pesticide detected was Roundup, with residues of glyphosate and AMPA in 9 of the 13 diets, up to 370 ppb. The levels correlated with the amount of Roundup-tolerant GMOs. Toxic effects of these pollutants on liver, neurodevelopment, and reproduction are documented. The sum of the hazard quotients of the pollutants in the diets (an estimator of risk with a threshold of 1) varied from 15.8 to 40.5. Thus the chronic consumption of these diets can be considered at risk. The high background rate of pathologies in laboratory rodents could be due to dietary contaminants. This invalidates the use of external controls (historical data) in regulatory tests, consisting of comparisons of toxicological effects to control rats from other experiments, because these control rats are fed different mixtures of pollutants. This also questions the use of 50 rats per group in carcinogenicity studies to increase the statistical power lost due to the elevated pathological background. Efforts toward safer diets will improve the reliability of toxicity tests in biomedical research.

W651: ONE HEALTH Epigenomics: From Soil to People
The Impact of GMOs and Glyphosate on Soil, Plant, Animal, and Human Health
Don Huber, Purdue University, TBD

W652: Organellar Genetics
Evolutionary Extremes in the Organelle Genomes of the Ancient Parasitic Plant *Hydnora visseri* (Hydnoraceae, Piperales)
Julia Naumann[1], Joshua P. Der[2], Eric Kenneth Wafula[1], Samuel S. Jones[1], Sarah T. Wagner[3], Gesine Schäfer[3], Christoph Neinhuis[3], Stefan Wanke[3] and Claude dePamphilis[1], (1)Penn State University, University Park, PA, (2)California State University, Fullerton, Fullerton, CA, (3)TU Dresden, Chair for Botany, Dresden, Germany
Plastid genomes of photosynthetic flowering plants are usually highly conserved in both structure and gene content. However, the plastomes of parasitic and mycoheterotrophic plants may be released from selective constraint due to the reduction or loss of photosynthetic ability. The nonphotosynthetic plant *Hydnora visseri* (Hydnoraceae, Piperales) is particularily interesting for several reasons, one is being among the oldest extant parasitic lineages.Genomic sequence data of *Hydnora* allowed us to identify and characterize its organellar genomes. The plastome is greatly reduced in size (only 27 kb in length), encoding 24 highly divergent genes. Several complementary approaches using the highest possible sensitivity were required for identification and annotation of this plastome. Despite the extreme reduction of the genome and high sequence divergence, the presence of syntenic, long transcriptionally-active open reading frames with distant similarity to other plastid genomes and a high plastome stoichiometry relative to the mitochondrial and nuclear genomes suggests that the plastome remains functional in *Hydnora*.In contrast to the plastome, the mitochondrial genome of *Hydnora visseri* is among the larger ones across the angiosperms (1.63 MB in length). The gene set is surprisingly conservative, representing a fairly complete and functional set of common plant mitochondrial genes.We demonstrate that divergent and reduced plastomes can now be detected in genomic NGS sequence data, where “traditional” methods required extensive effort. This greatly completes (and may change) our understanding of organellar genome evolution in nonphotosynthetic plants.

W653: Organellar Genetics
Progress in Implementing Plastid Transformation in *Arabidopsis thaliana*
Pal Maliga, Rutgers University, Piscataway, NJ
Arabidopsis thaliana, an important model plant species, is recalcitrant to plastid transformation. To enable early identification of transplastomic events, we developed a novel marker system that is selectively expressed in chloroplasts. We report on the successful deployment of the new marker system in *Arabidopsis* and discuss the requirements of developing a reproducible protocol. Plastid genome engineering in *Arabidopsis* will enable exploration of improving crop productivity by engineering the photosynthetic machinery, which would have a profound effect on basic science and applications in agriculture.

W654: Organellar Genetics
Genome-Wide Analysis of Translational Dynamics in Maize Chloroplasts
Alice Barkan and Prakitchai Chotewutmontri, University of Oregon, Eugene, OR
The translation of chloroplast mRNAs has long been known to be subject to regulation by developmental, environmental and physiological cues. However, progress in recognizing examples of translational regulation, identifying translational regulators, and dissecting mechanisms of translational modulation has been limited by the assays that have been available to monitor ribosome behavior *in vivo*: the traditional assays (pulse-labeling, polysome, and reporter gene approaches) are labor intensive, have limited sensitivity and/or resolution, and are not suited to genome-wide explorations. We are using genome-wide ribosome profiling methods that provide a quantitative and high resolution readout of ribosome positions *in vivo* to (i) identify nucleus-encoded proteins that are required for the translation of specific chloroplast mRNAs; (ii) analyze the impact of various light regimes on chloroplast ribosome behavior; (iii) describe the translational dynamics of chloroplast mRNAs during the differentiation of photosynthetic leaf cells. Examples of insights obtained in each of these areas will be discussed.

W655: Organellar Genetics
RNA Editing to Rescue Plants from Organelle Genome Mutations
Maureen Hanson, Cornell University, Ithaca, NY
Plant organelle genomes cope with T to C mutations by post-transcriptional correction of Cs to Us by RNA editing. The proper C must be selected for editing so as not to disrupt expression of correctly encoded codons. Individual members of the large nuclear-encoded pentatricopeptide repeat (PPR)-motif containing family are responsible for site-specific recognition of cis-elements 5’ to C targets of editing. Some members of this family contain C-terminal extensions termed the DYW domain that carry putative cytidine deaminase motifs

which, when mutated, disrupt RNA editing. The plant RNA editosome is a small RNA/protein complex less than 400 kD in size. Affinity purification followed by mass spectrometric analysis of protein content has led to the identification of three additional nuclear-encoded proteins required for editing in either chloroplasts or mitochondria: the RIP, ORRM, and OZ families. Mutations of genes encoding individual members of these families affect efficiency of editing of some, often many, but not all C targets. Thus, editosomes are diverse in composition, not only between different organelles, but within the same organelle. This flexibility in editosome constitution may have evolved in order to deal with new T to C mutations as they arose.

W656: Organellar Genetics

Insect Control by the Expression of Long Double-Stranded RNA in Plastids

Jiang Zhang, Hubei University, Wuhan, China

Double-stranded RNAs (dsRNAs) targeted against essential genes can trigger a lethal RNAi response in insect pests. However, although expression of dsRNAs targeted against insect genes in transgenic plants has impaired growth and development, complete protection of the plants and efficient killing of the insects were not achieved due to the presence of an endogenous RNAi pathway in plants that processes dsRNAs into short interfering RNAs (siRNAs). We found that long dsRNAs can be stably produced in plastids, a cellular compartment that appears to lack an RNAi machinery. When expressed from the chloroplast genome, dsRNAs accumulated to as much as 0.4% of the total cellular RNA. Transplastomic potato plants producing dsRNAs targeted against the β-actin gene of the Colorado potato beetle, a notorious agricultural pest, were efficiently protected from herbivory and were lethal to its larvae. Thus, plastid-expressed long dsRNAs can provide full crop protection without chemical pesticides and without synthesis of foreign proteins in the plant. Shifting the target of transgesis from the nucleus to the plastid removes the major hurdle on the way to exploiting transgenically delivered RNAi for efficient crop protection in the field.

W657: Ornamentals

Construction of Tetraploid Rose Maps and Analysis of Segregation Type

Paul Arens[1], Roeland E. Voorrips[2], Peter M Bourke[2], Tiago Santos Leonardo[3], Danny GE Esselink[1], Mirjana Vukosavljev[4], Carole FS Koning-Boucoiran[5], Chris Maliepaard[2] and Marinus J. M. Smulders[6], (1)Wageningen UR, Wageningen, Netherlands, (2)Wageningen UR Plant Breeding, Wageningen, Netherlands, (3)HZPC, Joure, Netherlands, (4)Pheno Geno, Novi Sad, Serbia, (5)HAN, Nijmegen, Netherlands, (6)Wageningen UR - Plant Breeding, Wageningen, Netherlands

Polyploidy is common in many important crop plants and especially in ornamentals. In polyploids more than two different alleles may occur per locus in one plant, while any given allele may exist in the plant in multiple ‘dosages’. This puts constrains on genetic analyses in polyploids. To resolve these constrains one can start with constructing separate maps for each haplogroup, but this means that many hundreds to thousands of markers are needed to construct genetic maps. Such high numbers of SNP markers can nowadays be identified using next generation sequencing, and genotyped on SNP arrays.

We developed a 68k SNP Axiom array (WagRhSNP) for rose based on transcriptome data and used them to genotype tetraploid garden and cut rose F1 populations. SNPs were scored using software for dosage scoring (fitTetra, Voorrips et al., BMC Bioinform 12:172, 2011). Results of mapping, synteny of the maps to Fragaria, and analyses of the segregation type of rose (random vs preferential pairing) will be presented.

W658: Ornamentals

Transcriptome Analysis of the Defense Response of Roses in the Interaction with *Diplocarpon rosae* and *Podosphaera pannosa* using the MACE Technique

Enzo Klein, Helgard Kaufmann and Thomas Debener, Leibniz University of Hannover, Hannover, Germany

Roses are the economically most important ornamental plants worldwide. Besides the research on horticultural important traits like flower color or scent, the defense response against pathogens comes into focus more and more.

The presented work deals with the defense responses of roses against two major fungal pathogens *Diplocarpon rosae* and *Podosphaera pannosa*. These interaction systems are of particular interest because of the different lifestyles of the pathogens. *D. rosae* is a hembiotrophic ascomycet, whereas *P. pannosa*is an obligate biotrophic ascomycet.

We performed a transcriptome analysis with the new MACE (Massive Analysis of cDNA Ends) technique to obtain data for three interaction systems: a compatible interaction of the susceptible rose variety “Pariser Charme” inoculated with *P. pannosa*, a compatible interaction of “Pariser Charme” with *D. rosae* and an incompatible interaction of the resistant genotype 91/100-5 with *D. rosae*.

Non-infected “Pariser Charme” leaves were used as controls. Samples for three different time points (0, 24, 72 hpi) were taken and all inoculation experiments have been performed three times independently. All in all, we sequenced 33 libraries and obtained an average of 12.7 million reads per library where the read count ranged from 6.2 to 32.3 million reads.

The first results indicate that typical defense responses are visible in all interaction systems. We could also identify sequences which are specifically induced as response to either one of the pathogens or to the resistance reaction in the incompatible interaction. Additionally, we could isolate fungal sequences of both pathogens.

W659: Ornamentals

The Rose Genome Sequencing Initiative, Prospects and Perspectives

Mohammed Bendahmane[1], J. Just[1], P. Vergne[1], O. Raymond[1], A. Dubois[1], J. Szcesi[1], S-H. Yang[1,2], Xiaopeng Fu[1,3], J Gouzy[4], S. Carrere[4], L. Legrand[4], P. Wincker[5], A. Lemainque[5], J-M. Aury[5], A. Couloux[5], F. Foucher[6], L. Hamama[6], L. Hibrand-Saint-Oyant[6], S. Sakr[6], N. Choisne[7], H. Quesneville[7], M. Le Bris[8], B. Goveto[8], S. Baudino[9], S. Moja[9], JC Caissard[9], K. Van laere[10], J De Riek[10], T. Debener[11], Mjm Smulders[12], J. Salse[13], D. Zamir[14], D. Sargent[15], A. Bruneau[16], Huijun Yan[17], Kaixue Tang[17], D. Byrne[18], S. Hokanson[19], A. Torres[20], I Amaya[20] and M. Benhamed[21], (1)Ecole Normale Supérieure, Lyon, France, (2)Chinese Academy of Agricultural Science, Beijing, China, (3)Huazhong Agricultural University, Wuhan, China, (4)LIPM-INRA Toulouse, Castanet-Tolosan, France, (5)CEA-IG-Genoscope-CNS, Evry, France, (6)INRA Angers, Beaucouzé, France, (7)URGI, INRA Versailles,

Versailles, France, (8)Université Aix Marseille, Marseille, France, (9)Université Jean Monnet, St Etienne, France, (10)ILVO, Melle, Belgium, (11)Leibnitz University Hannover, Hannover, Germany, (12)Wageningen UR Plant Breeding, Wageningen, Netherlands, (13)INRA Clermont-Crouel, Clermont-Ferrand, France, (14)Hebrew University of Jerusalem, Israel, (15)Fondazione Edmund Mach, S. Michele all'Adige, Italy, (16)Montreal Botanical Garden, Montreal, QC, Canada, (17)Yunnan Academy of Agricultural Sciences, Kunming, China, (18)Texas A&M University, College Station, TX, (19)University of Minessota, Saint Paul, MN, (20)IFAPA, Malaga, Spain, (21)Paris-Sud University, Orsay, France
Roses are of high symbolic value and have great cultural importance in different societies worldwide. The rose is well suited to be an original model organism for woody ornamental species as it has a relatively small genome size (560Mbp) and it has a short life cycle for a perennial woody plant. Several characters, such as recurrent blooming, flower morphogenesis, scent(1),… are of high economic importance. During the past years, we generated a number of molecular, genomic and biotechnology tools(2) such as efficient genetic transformation(3) and a database that provides useful information on Rosa expressed genes with thorough annotation and an overview of gene expression patterns with good accuracy(4) ; the latest represented a valuable prerequisite to the the rose genome sequencing. We have undertaken the genome sequencing of the diploid *Rosa chinensis* cv Old Blush, an an ancestor of modern roses that contributed several important characters. We generated and assembled a high quality genome sequence for this cultivar, but its relatively high heterozygosity hampered high quality genome assembly. To overcome this difficulty, we successfully generated a *R. chinensis* rose homozygous tissue (*HzRc-RDP12*) using Old Blush as starting material. The sequencing and assembly of this homozygous genotype is in progress. The availability of the rose genome sequence will be of great help for discovering the molecular and genetic mechanisms controlling many horticultural traits in *Rosa sp.* and likely in other ornamental species.
(1)Magnard 2015 Science 349:81-3. (2)Bendahmane 2014 J Exp Bot 64:847-57. (3)Vergne 2010. PCTOC 100:73-81. (4)Dubois 2012 BMCGenomics 13:638.

W660: Ornamentals

French Rose (Rosa L. sp) Germplasm from the 19th Century Shows a Continuous Shift throughout the Time from a European to an Asian Genetic Background

Mathilde Liorzou[1], Alix Pernet[1], Shubin Li[1,2], Annie Chastellier[1], Tatiana Thouroude[1], Gilles Michel[1], Valéry Malécot[1], Sylvain Gaillard[1], Céline Briée[3], Fabrice Foucher[1], Cristiana Oghina-Pavie[3], Jérémy Clotault[1] and Agnès Grapin[1], (1)Research Institute of Horticulture and Seeds (IRHS), Beaucouzé, France, (2)Yunnan Academy of Agricultural Sciences, Kunming, China, (3)CERHIO, Centre de recherches historiques de l'Ouest, Angers, France
Innovation in ornamental plant breeding is commonly obtained by hybridization with newly introduced genetic resources. In France, the 19th century was a golden age for rose breeding and this period corresponds to the introduction and subsequent use for breeding of Asian roses in Europe. Our objective here was to study and explain the evolution of roses genetic diversity during this period of time. A large sample of 1,228 garden roses sampled in ten French rose gardens was constituted. This sample summarized the conserved diversity of roses cultivated in Europe during the 18th and 19th centuries as it included 991 European garden roses from this period of time. As a comparison, Asian, botanical and modern roses were included to the study. Genotypes of various ploidy levels, ranging within 2 to 6 x, were found in the sample. The individuals were genotyped with 32 single sequence repeat (SSR) primer pairs and the genetic diversity and structure of the sample were assessed. A wide diversity, structured into sixteen genetic groups, was observed. A genetic differentiation was detected between ancient European and Asian accessions and a continuous temporal shift was observed in cultivated hybrids from a European to an Asian genetic background during the 19th century. Frequent crosses with Asian roses along the 19th century and/or selection for Asiatic traits may have induced this shift. Implications for rose breeding and germplasm conservation will be discussed.

W661: Ornamentals

Mapping Black Spot Resistance in Autotetraploid Rose Using Genotyping-by-Sequencing

Travis W. Banks, Vineland Research and Innovation Centre, Vineland Station, ON, Canada
The black spot pathogen, *Diplocarpon rosae*, creates significant damage to landscape roses. Once infection has occurred the leaves of affected plants become colonized by unsightly black patches which can be followed by significant defoliation and the eventual death of the plant. Control of this endemic pathogen is difficult particularly as industry phases-out the use of fungicides. Rose breeding programs around the world could benefit from molecular markers for black spot resistance but the complex history and genetics of rose combined with under-developed genomics resources has limited efforts.

Here we present our work to map two races of black spot in autotetraploid rose. Using genotyping-by-sequencing (GBS) we constructed parental maps from 350 hybrids from a cross between accession 564 and Gentle Giant®. Using a detached leaf assay we phenotyped resistance to both races, each of which mapped as a single dominant gene. The GBS SNPs were converted to high resolution DNA melting assays and have been deployed to our cold hardy landscape rose breeding program. This project will serve as a framework for future development of markers across a number of traits important for developing superior roses.

W662: Ornamentals

RNAi Suppression of Two *AGAMOUS* Homologs in Sweetgum (*Liquidambar*) Impairs Male and Female Reproductive Development Under Field Conditions

Amy Leigh Klocko, Kori Ault, Cathleen Ma and Steven Strauss, Oregon State University, Corvallis, OR
Many woody ornamentals are grown as exotics outside of their native range, where they have the potential to become invasive. As street trees, their fruits often create sanitation problems and their pollen exacerbates allergies. A reduction in fertility of these ornamental plants would help to mitigate these problems. We used RNA interference (RNAi) to suppress the expression of a key floral development gene, *AGAMOUS* (*AG*) in sweetgum trees. Sweetgum is a popular ornamental tree in the United States known for its vibrant fall foliage, however, it is also known for producing copious amounts of hard, rough fruits termed gum balls or burr balls. Shed fruits are long-lasting and are an unwelcome nuisance on streets, roofs, and often clog gutters. We used *Agrobacterium* to transform sweetgum variety 'Worplesdon' with a double-stranded RNAi-inducing transgene that targets the two distinct *AG* orthologs in sweetgum. A total of 33 independent transgenic events, plus 12 non-transgenic

controls, were planted in the field in 2006. The RNAi-*AG* sweetgum trees maintained normal growth, phenology, and vivid fall coloration during 9 years of study. We found that 8 events had highly-modified floral morphology, which failed to produce seeds or pollen, and had anthers and carpels that were converted to flat leaf-like structures with no ovules. The female flowers from these events developed into dry papery fruits lacking seeds. All of these traits were stable across multiple growing seasons. RNAi against *AG* is highly effective at modifying fertility and reproductive development in sweetgum.

W663: Ornamentals

Map Construction in Diploid Rose with GBS

Muqing Yan, Texas A&M University, college station, TX

Black spot disease (*Diplocarpon rosae* (Lib.) Wolf) of rose is the most important leaf disease of garden roses in warm humid areas. Although the partial (horizontal) resistance to black spot has been shown to be moderately heritable, the responsible quantitative trait loci (QTLs) remain unidentified. Because of the interspecific nature and high heterozygosity in commercial roses the genetic information available on rose is limited. To effectively identify markers associated with the QTL(s) controlling black spot resistance, one needs abundant markers along the genome and careful phenotyping. The objective of this study is to utilize genotyping by sequencing technology to generate thousands of informative single nucleotide polymorphism (SNP) markers for genetic linkage and QTL mapping. Thus far rose seedlings and parents have been phenotyped for partial black spot resistance in the lab and the field, and an efficient protocol to extract high quality DNA for sequencing from rose leaves has been developed. Seven diploid rose populations created from the crosses of black spot resistant breeding lines derived from *R. wichuriana* 'Basye's Thornless' with susceptible commercial cultivars ('Vineyard Song', 'Red Fairy', 'Sweet Chariot' and 'Little Chief') were used for SNP detection. Their genomic DNA was digested using methylation sensitive enzymes and the resulting fragments were sequenced using the Illumina Hiseq 2500 platform. Based on preliminary data, about 89-91% of the sequencing reads were aligned to the rose contigs and around 50% of the reads were mapped to the strawberry genome, respectively. The SNP markers detected based on strawberry genome along with SSR (simple sequence repeat) data will be utilized together to create a diploid consensus map and identify markers associated with the partial black spot resistance trait.

W664: Palm Genetics and Genomics

Genomic Strategies for Oil Palm Improvement

Rajinder Singh[1], Eng-Ti L Low[1], Meilina Ong Abdullah[1], Rajanaidu Nookiah[1], Mohamad Arif Abdul Manaf[1], Muhammad Arief Budiman[2], Nathan Lakey[2], Blaire Bacher[2], Andrew Van Brunt[2], Steven W Smith[2], Jared Ordway[2], Rob Martienssen[3] and Ravigadevi Sambanthamurthi[1], (1)Malaysian Palm Oil Board, Kajang, Malaysia, (2)Orion Genomics LLC, St. Louis, MO, (3)HHMI-GBMF Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The oil palm with its long breeding cycle and large land requirement for field trials is an ideal candidate for genomics guided breeding. However, oil palm improvement even until recently was mostly due to conventional breeding. This is because the linking of markers and gene(s) to traits of interest has been slow for this perennial crop, partly due to the lack of in-depth sequence information in the past. A modest effort to start sequencing the genome was thus initiated in 2004, which led to sequencing of the hypomethylated regions of the genome. This initiative provided the impetus to sequence the whole genome, which was successfully executed in 2013. The early focus after sequencing the genome was to identify genes influencing important monogenically inherited traits. The availability of a large germplasm collection and a well-executed breeding programme allowed the team to identify the genes influencing fruit form and colour of oil palm. The early discoveries has formed the basis for marker assisted selection in oil palm as well as provided the community tools for quality control in commercial seed production. The focus has now shifted to more complex traits, where stable quantitative trait loci (QTLs) linked to fatty acid composition have been identified. MPOB and its consortium partners have also unraveled the epigenome of oil palm to help understand DNA methylation alterations in clonal palms. Deciphering of the epigenome and understanding the causes of clonal abnormality has also made large scale tissue culture of oil palm feasible.

W665: Palm Genetics and Genomics

Experiences in Assembling and Annotating the Oil Palm Genome

Eng-Ti L Low[1], Kuang-Lim Chan[1], Mohd Amin Halim[1], Corey Wischmeyer[2], Steven W Smith[2], Rozana Rosli[1], Norazah Azizi[1], Nik Shazana Nik Mohd Sanusi[1], Nadzirah Amiruddin[1], Jayanthi Nagappan[1], Rajinder Singh[1], Meilina Ong Abdullah[1], Rajanaidu Nookiah[1], Tatiana Tatarinova[3], Petr Ponomarenko[3], Martin Triska[4], Leslie Cheng-Li Ooi[1], Pek Lan Chan[1], Ngoot-Chin Ting[1], Michael Hogan[2], Blaire Bacher[2], Muhammad Arief Budiman[2], Jared Ordway[2], Nathan Lakey[2], Rob Martienssen[5] and Ravigadevi Sambanthamurthi[1], (1)Malaysian Palm Oil Board, Kajang, Malaysia, (2)Orion Genomics LLC, St. Louis, MO, (3)University of Southern California, Los Angeles, CA, (4)Children's Hospital Los Angeles, Los Angeles, CA, (5)HHMI-GBMF Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

The oil palm is the most productive oil crop, producing an average of ~4 tonnes of oil per hectare per year in Malaysia. However, the full potential of the crop has not been realized as individual palms in breeding trials have been shown to have yields as high as 14 tonnes of oil per hectare per year. The oil palm has a long breeding cycle (8-12 years), which makes the introduction of new commercial elite planting materials time consuming. This however, can be expedited through the use of genomics-guided breeding, thus leading us to initiate and develop an oil palm genome programme. The oil palm genome, which is approximately 1.8Gb was sequenced, assembled, annotated and published in 2013 using a combination of NGS paired-end and fragment libraries, as well as BAC end sequencing with the Sanger technique. Approximately 44% of the assembled genome was incorporated into pseudochromosomes using genetic maps. Since then, the genome has been significantly improved and updated to allow researchers to use the data more efficiently. This was achieved using new sequencing technologies and by improving the assembly and annotation process. An oil palm genome browser (MYPalmViewer) has also been developed. The browser can be accessed via a link in the Genomsawit portal (http://genomsawit.mpob.gov.my) or directly via http://gbrowse.mpob.gov.my. Here we provide an overview of the developments in MPOB's Oil Palm Genome Programme and experiences in assembling and annotating the oil palm genome.

W666: Palm Genetics and Genomics

Application of Genotyping-by-Sequencing (GBS) for SNP-Based Linkage Mapconstruction and Identification of QTL Associated with Trunk Height in Oil Palm

Wirulda Pootakham, National Center for Genetic Engineering and Biotechnology, Pathum Thani, Thailand

Oil palm (*Elaeis guineensis*) has become one of the most important oil crops in the world. Marker-assisted selections have played a pivotal role in oil palm breeding programs as they reduce the amount of time and resources required to develop new cultivars. Single nucleotide polymorphisms (SNPs) have recently become markers of choice owing to their abundance in the genome. We applied a genotyping-by-sequencing (GBS) approach for a large-scale SNP discovery and genotyping of a mapping population. We subsequently used the SNP-based linkage map for the identification of quantitative trait loci (QTL) associated with trunk height. Reduced representation libraries of 108 F_2 progeny were sequenced and a total of 524 million reads covering 56 Gb of high-quality data were obtained. We detected 21,471 single nucleotide substitutions, most of which represented transition events. Of 3,417 fully informative SNP markers, we were able to place 1,085 on a linkage map, with an average of 64 SNPs per linkage group. The map spanned 1,429 cM and had an average of one marker every 1.26 cM. Three QTL affecting trunk height were detected on LG 10, 14 and 15. Interestingly, the QTL governing stem stature identified on LG 14 was linked to two open reading frames encoding a putative gibberellin 2-oxidase and a putative DELLA protein GAI1. Both proteins have been implicated in plant height regulation via gibberellin homeostasis and signaling pathway. Finally, two SNP markers associated with the major QTL were validated using the TaqMan assays in the population with similar genetic background.

W667: Palm Genetics and Genomics

Empirical Prediction Accuracy of Genomic Selection Between Experimental Designs and Generations in Oil Palm

David Cros, CIRAD, UMR AGAP, Montpellier, France

There is a large potential for the genetic improvement of yield in oil palm, as it has only been submitted to a few generations of modern breeding. Selection candidates are traditionally evaluated in progeny tests, as some yield components have a low heritability. The progeny-tests allow selecting with high accuracy, but constrain the rate of genetic gain by increasing the generation interval and limiting the number of evaluated individuals. Genomic selection (GS) is an appealing alternative that could allow selecting without progeny tests.

We studied the prediction accuracy of GS, ie the correlation between the genomic estimated breeding values and the breeding values obtained from the conventional selection method, in the two heterotic groups used to produce commercial hybrids. We used two independent experimental designs located in Sumatra (Indonesia), comprising over 700 hybrid crosses. The GS model was calibrated with the first experimental design (training set) and applied to predict the breeding values of the individuals progeny tested in the second design (validation set). The genomic breeding values were obtained by GBLUP, using the phenotypes of the training hybrids and the genotypes of all the progeny tested individuals. The genotypes consisted in 3000+ SNPs from genotyping-by-sequencing (GBS).

We found that it was possible to achieve satisfactory GS accuracy provided the set of SNPs was carefully defined (density and filtering criteria). Finally, GS appeared as a highly valuable approach to speed up the genetic progress in oil palm yield, and GBS as a suitable technology to generate the genotypes required.

W668: Palm Genetics and Genomics

Sequence Capture Methods Help Resolve Relationships Among the Arecaceae

Karolina Heyduk[1], Jason R. Comer[1], Christine D. Bacon[2] and Jim Leebens-Mack[1], (1)University of Georgia, Athens, GA, (2)University of Gothenburg, Göteborg, Sweden

Understanding evolutionary relationships among taxa is crucial for the study of trait evolution and comparative biology, both of which can inform genomics and breeding research. A number of processes, including low substitution rates, hybridization, and retention of ancestral polymorphism, can complicate our estimates of species trees. By generating a species tree based on a larger number of genes, these problems may be minimized. Here we describe the use of sequence capture to selectively sequence plastid and low-copy nuclear genes from across populations, species, and tribes in the Arecaceae. Biotinylated RNA baits are designed complementary to exons, and when hybridized to the target DNA, can be pulled from solution with by the biotin label and subsequently sequenced. Additionally, the adjacent intron sequence is captured, resulting in a combination of conserved sites from the exon and phylogenetically informative sites from the intron. A single bait set has been designed for nuclear and chloroplast sequences in the Arecaceae, and has worked at broad taxonomic levels. Future work could utilize the same capture baits or bait design strategy to address a wide variety of questions concerning evolutionary relationships or population structure across a broad diversity of palm lineages.

W669: Perennial Grasses

Transcriptome Analysis of Tall Fescue Endophyte-Infected and Endophyte-Free Tissues Under Stress Conditions

Randy Dinkins, USDA-ARS, Forage-Animal Production Research Unit, Lexington, KY

Tall fescue (*Lolium arundinaceum*) plants symbiotic with the endophytic fungus, *Epichloe coenophiala* , (E+), have been shown to have better survivability and persistence than plants lacking the endophyte (E-). To understand more about the grass-endophyte interactions and how endophyte affects the host plant physiology and gene expression, a greenhouse time course study was conducted in which water was withheld from 0 - 5 days, and endophyte effects were investigated on plant and fungal metabolites and gene expression profiles at the two day water-withholding time point. For gene expression analysis, an in house tall fescue transcriptome assembly was generated using the RNA-Seq reads and mapping onto the assembly was done using CLC Genomics Workbench. Analysis of differential gene expression was done using JMP Genomics and significant differences (>4-fold; $P<0.001$) was calculated using computed adjustments for each contrast. Comparison of whole transcriptome expression profiles of the control versus stressed plants revealed a large number genes differentially expressed due to the stress treatment, and is similar to what has been shown previously for number of other species. When comparing the differentially expressed gene between the E+ and E- plants under stress conditions, a lower number of differentially expressed were observed in the E+ tissues than in the E- tissues suggesting that the presence of the endophyte may modulate the effect of the stress imposed. Results of the pathways affected will be presented and discussed.

W670: Perennial Grasses

Genomic Progress in Bermudgrass

Patricio R. Munoz[1], Jose Dubeux[2], William Anderson[3], Joao Vendramini[4], Brian Shwartz[5], Miguel Castillo[6], Susana Milla-Lewis[6] and Alexandra Rucker[7], (1)University of Florida, Agronomy Department, Gainesville, FL, (2)University of Florida, Marianna, FL, (3)USDA, Tifton, GA, (4)University of Florida, Ona, FL, (5)University of Georgia, Tifton, GA, (6)North Carolina State University, Raleigh, NC, (7)University of Florida, Gainesville, FL

Bermudagrass (*Cynodon spp*) has been grown extensively in the southeastern USA for more than 50 years as the preferred warm-season perennial grass for forage and livestock producers. Producers prefer this grass because of its pereenial growth habit, high yield, long growing season (April-October), persistence under grazing, and relatively high forage quality. In the past, bermudagrass breeding has focused on yield, quality and grazing persistence; however, the most popular cultivars were developed more than 40 years ago. In a collaborative effort involving five research institutions (University of Florida, North Carolina State University, USDA-Tifton, University of Georgia, and Noble Foundation) and seven experimental sites, a clonal association population (n~280) has been established representing most environmental conditions of the southern United States. The objectives of this study are: 1) screen a large collection of individuals to select parental lines, and 2) to find QTLs for economically important traits, such as bermudagrass stem maggot (*Atherigona reversura*) and nitrogen concentration. The experiments were established in summer 2014 and 2015 with two repetitions per site in a row-column design. A large volume of phenotypic data has been recorded since establishment with an emphasis on yield, quality, leaf N concentration and pest/diseases. Genotyping of the population using the genotype-by-sequence method is in progress. Results from the GWAS analysis will be presented and discussed.

W671: Perennial Grasses

Genome-Wide Divergence Between Upland and Lowland Ecotypes of *Panicum hallii*, a Close Relative of Switchgrass

John T. Lovell, University of Texas, Austin, Austin, TX

Local adaptation to abiotic stress is responsible for much of the observed genetic diversity within wild plant species and landraces of crops. For example, selection on drought adaptation strategies can contribute to the formation of upland (drought tolerant) and lowland (mesic) ecotypes. Such ecotypes have have diverged at many thousands of loci and most measurable phenotypes. Interestingly, when crossed, upland-lowland hybrids often exhibit heterosis, indicating that deleterious mutations have accumulated alongside adaptive sequence variants. Here, we characterize sequence and gene expression divergence and analyze sequence variants that may underlie whole-plant heterosis between upland and lowland ecotypes of *Panicum hallii*, a diploid genomic model of Switchgrass.

W672: Perennial Grasses

Integrated Transcriptomic and Epigenomic Analysis in Switchgrass

Venu (Kal) Kalavacharla, Molecular Genetics & Epigenomics Laboratory, Delaware State University, Dover, DE

Epigenetic mechanisms regulate several physiological processes including plant growth, development, signaling, and abiotic and biotic stress-responses. Switchgrass is a dedicated herbaceous bioenergy feedstock for cellulosic ethanol production. The main objective in switchgrass research has been the improvement of biomass yield, production and its lignocellulosic conversion into biofuel. Until now, genomic and transcriptomic approaches have been utilized in addressing yield-related traits but integrated epigenomic and transcriptomic analyses have not been undertaken, especially the role of chromatin modifications on gene expression. Genome-wide transcriptome profiling revealed 6,619 and 5,369 genes that were differentially expressed between upland (AP13) and lowland (VS16) ecotypes, respectively. Gene ontology and KEGG pathway annotations identified genes involved in phenylpropanoid, C4 and photorespiration-related pathways, and drought-, heat-, and salinity-tolerance, transporters in stress, and disease-resistance in the two ecotypes. The majority of pathways identified here belonged to metabolic processes. To understand chromatin modification at the whole-genome level, we used chromatin immunoprecipitation sequencing (ChIP-Seq) to analyze DNA binding of $H3K9_{me2}$ and $H4K12_{ac}$ marks in both genotypes. More than 70% of the sequenced reads were mapped to reference genome (AP13). The number of histone-DNA binding regions was higher in $H4K12_{ac}$ than $H4K9_{me2}$ in both genotypes. About 60% of the peaks were related to genic region (upstream, 5'-UTR, exon, intron, 3'-UTR and downstream) in both modifications. This integrated analysis will aid not only in developing reference epigenomes and transcriptomes in switchgrass but also in understanding the interplay between epigenomic modulators and gene expression in plant stress responses.

W673: Perennial Grasses

Expected and Unexpected Patterns of Chromosomal Inheritance from Resequencing of Tetraploid Switchgrass

Laura Bartley[1], G. Albert Wu[2], Yanqi Wu[3], Daniel S. Rokhsar[2], Jeremy Schmutz[4], Malay C. Saha[5], Kerrie W. Barry[2], Sandra Thibivilliers[1], Thomas Juenger[6], David Lowry[7], C. Robin Buell[8], Joseph Evans[8], Michael Casler[9], Carol Auer[10], Katrien M. Devos[11] and E. Charles Brummer[12], (1)University of Oklahoma, Norman, OK, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)Oklahoma State University, Stillwater, OK, (4)Hudson Alpha, Huntsville, AL, (5)The Samuel Roberts Noble Foundation, Ardmore, OK, (6)University of Texas, Austin, TX, (7)Michigan State University, East Lansing, MI, (8)Department of Plant Biology and DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing, MI, (9)USDA Dairy Forage Research Center, Madison, WI, (10)University of Connecticut, Storrs, CT, (11)Institute of Plant Breeding, Genetics and Genomics, and Dept. of Plant Biology, University of Georgia, Athens, GA, (12)University of California, Davis, Davis, CA

Switchgrass (*Panicum virgatum* L.) is a large-stature, stress-tolerant North American Tallgrass Prairie species being developed as a bioenergy crop. It is primarily outcrossing, consists of upland and lowland ecotypes, and is typically polyploid. The main objectives of this project are to produce 1) sequence of reduced diversity genotypes to support the switchgrass genome project and 2) a high-quality, uniform variant catalog to support switchgrass improvement efforts. In this project, 48 genotypes of switchgrass have been resequenced via shotgun sequencing, with ~60 Gbp of 150 bp paired-end reads generated per genotype. Resequenced genotypes consist of lowland tetraploids (33), uplands (11, including 4 octoploids), and lowland-upland hybrids (4). Two of the lowlands possess upland chloroplast sequences, consistent with hybridization across the

ecotypes *in situ*. Sequencing confirmed that three genotypes have greatly reduced heterozygosity relative to the reference, including a dihaploid, and two third-generation selfed genotypes. The data show approximately the expected crossover frequency upon selfing. Preliminary analysis has found that many of the genotypes, including some wild accessions, possess large regions that are homozygous between the sister chromatids. Low heterozygosity regions are also apparent in other switchgrass genotypes characterized by exome capture. This suggests the possibility of 1) rampant nondisjunction during chromosomal segregation, 2) the opportunity to study/isolate homozygous alleles in these regions, and 3) improving breeding program diversity by incorporating genotypes with variation in these regions. The data provide a deep characterization of the genetic differentiation between individuals and within ecotypes that might be leveraged to improve switchgrass as a bioenergy species.

W674: Perennial Grasses

Application of Switchgrass Genetics to Biocontainment and Conservation

Carol Auer, University of Connecticut, Storrs, CT

Switchgrass (*Panicum virgatum*) is a native grass developed for bioenergy, but questions exist about the long-term impacts of gene flow. This presentation reports new information about switchgrass distribution, genetic diversity, pollen dispersal, and the use of forest windbreaks to mitigate gene flow. In the first project, SSR markers were used to determine the genetic relationship between 122 plants collected in the Northeastern US and 28 reference cultivars. Results showed that 54% of the coastal plants were a unique Lowland tetraploid genotype. Most roadside plants (67%) were Upland octoploids introduced by humans and adapted to disturbed habitats. The second project modeled switchgrass pollen dispersal under two different wind patterns. Switchgrass pollen entrained in light wind conditions with buoyancy-driven turbulence moved up to 3.5 km from the source field. Pollen carried in pressure-driven, non-turbulent wind conditions moved up to 6.5 km. However, most pollen was deposited close to the source field. The third experiment measured the ability of a forest windbreak to mitigate pollen drift. The project produced the first estimate of switchgrass pollen source strength at 141 billion pollen/hectare/year. The forest windbreak consistently decreased downwind switchgrass pollen concentrations by 333-20,000 fold compared to the control plot which showed a much smaller decrease due to downwind distance alone. Thus, forest windbreaks could be designed to decrease gene flow while providing ecosystem services. Together, these experiments support ecological risk assessments, predictions about pollen-mediated gene flow, coexistence strategies, and the conservation of genetic diversity in switchgrass. Supported by USDA NIFA Biotechnology Risk Assessment grants program.

W675: Plant Chromosome Biology

Evolution of Reproductive Barriers in Tomatoes

Yongbiao Xue, Institute of Genomics, Beijing, China

TBA

W676: Plant Chromosome Biology

Directed Chromosome Engineering *via* Uniparental Genome Elimination

Ek Han Tan[1], Anne B. Britt[1], Awais Khan[2], Merideth Bonierbale[2] and Luca Comai[1], (1)University of California, Davis, CA, (2)International Potato Center, Lima, Peru

Plant breeding through haploid production is an efficient way to fix traits of interest. Haploid production is possible in some species, but problematic in many. In *Arabidopsis thaliana*, genome elimination and haploidy can be induced by modification of centromeric histone H3, an approach that should be applicable in other species. We plan to test this method in cultivated potato (*Solanum tuberosum* Group Tuberosum) and compare it to haploid induction resulting from the natural inducer, *Solanum tuberosum* Group Phureja. Secondly, DNA from the haploid inducer can sometimes be retained and inherited during genome elimination in Arabidopsis, effectively providing an unusual case of genomic introgression. A related process has been suggested by genotyping Phujera induced haploids in potato. We aim to compare the features of these systems to exploit novel ways in which haploid induction could facilitate breeding of crop plants.

W677: Plant Chromosome Biology

Proteomic Analysis of Plant Mitotic Chromosomes

Beata Petrovska[1], Hana Jerabkova[1], Jana Beinhauer[2], Ivo Chamrad[2], Jan Vrana[1], Rene Lenobel[2], Jana Urinovska[2], Nicolas Blavet[1], Marek Sebela[2] and **Jaroslav Dolezel**[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)Palacky University in Olomouc, Olomouc, Czech Republic

Transmission of hereditary information to daughter cells depends critically on the formation of chromosomes and their ordered behavior. Condensation of nuclear chromosome domains into compact mitotic chromosomes involves interaction of DNA with histone and non-histone proteins, some of which are modified during the process. Chromosome condensation is accompanied by formation of proteinaceous kinetochore domain, which binds to mitotic spindle, controls separation of sister chromatids and their movement towards mitotic spindle poles. Despite their obvious significance, apart for small basic histone proteins, the knowledge of chromosome proteins in plants remains limited. As the first step towards obtaining a more complete picture of plant chromosome protein composition, we performed proteomic analysis of barley mitotic metaphase chromosomes. The work involved chromosome purification by flow cytometric sorting followed by protein extraction, separation and digestion with the subsequent mass spectrometry of purified peptides. This approach allowed identification of a large number of proteins that form mitotic chromosomes. The protocol does not affect chromosome protein composition and minimizes the risk of contamination by non-chromosomal (e.g. cytoplasmic remnants) proteins. Thus, the approach should contribute towards understanding the macromolecular organization of plant mitotic chromosomes and revealing the processes that accompany their formation and influence their behavior. Subsequent work will focus on specific roles of each of the newly identified proteins in chromosome structure and function. This research was supported by grant awards from the Czech Science Foundation (14-28443S) and the National Program of Sustainability I (LO1204) from the Ministry of Education, Youth and Sports of the Czech Republic.

W678: Plant Chromosome Biology

New Insights into Meiotic Recombination from Analyses of DNA Synthesis Gene Functions

Yingxiang Wang, State Key Lab of Genet Eng, Sch of Life Sci., Fudan Univ., Shanghai, China
Meiosis halves diploid genomes to haploid and is essential for sexual reproduction in eukaryotes. Meiotic recombination ensures the proper segregation of homologous chromosomes and also results in the redistribution of genetic materials among progeny. DNA synthesis is essential for meiotic recombination, but the molecular mechanisms of DNA synthesis in meiotic recombination are largely unknown. We previously used RNA-seq to investigate the *Arabidopsis* male meiocytes and found many DNA synthesis genes with higher expression at meiocytes, suggesting their potential roles in meiosis. Further studies using molecular genetic, cell biological and biochemical approaches provided strong evidence that lagging strand DNA synthesis genes are required for the formation of meiotic interference-sensitive crossovers (COs), supporting the idea that meiotic recombination requires lagging strand DNA synthesis. Moreover, genome re-sequencing analyses in both yeast and plants showed that meiotic gene conversion tracts vary in length, ranging from several hundreds to thousands of nucleotides and that tracts associated with CO are usually longer than those of NCO. This raises a question whether there is any difference in DNA synthesis among CO-associated tracts with different lengths. Analyses of the largest subunit of DNA polymerase ε, important for leading strand elongation, demonstrated that it has a role in promoting formation of a portion of meiotic interference-sensitive COs, suggesting that sufficient leading strand DNA elongation could be a critical step of interference-sensitive/insensitive CO differentiation. Together, given that meiotic recombination and DNA synthesis are highly conserved in divergent eukaryotes, our studies suggest a novel role for DNA synthesis in the differentiation of meiotic recombination pathways.

W679: Plant Chromosome Biology

Evolution of Centromeres in Common Bean and Cowpea

Aiko Otsubo and Scott A. Jackson, University of Georgia, Athens, GA
In higher eukaryotes, centromeres typically contain large arrays of satellite repeats that evolve rapidly and homogenize within a genome resulting in species-specific centromeric repeats. It is unclear why centromeric repeats evolving rapidly while function is conserved. In addition, our knowledge of the structure and evolution of centromeric satellites within a genome is limited. Here, we report the analysis of centromeric repeats of two legume species, *Phaseolus vulgaris* (common bean) and *Vigna unguiculata* (cowpea).
In common bean, two satellite repeats are present at centromeres; CentPv1 on eight centromeres and CentPv2 on three centromeres. Using a genomics and cytogenetic approach, we found several chromosome-specific variants of CentPv1 and CentPv2 that originated after the divergence of common bean from other *Phaseolus* species. Centromeric satellite repeats and other repeats were used to develop a fluorescence *in situ* hybridization (FISH)-based karyotype map for common bean. This karyotype map allowed us to further analyze centromere evolution within common bean and other related *Phaseolus* species. Comparison of karyotype maps within closely related species indicated that chromosomal distribution of the centromeric satellite repeats are stable, while the copy number of the repeats was subject to change. In other more diverged *Phaseolus* species, ~2-4 million years, copy numbers of CentPv1 and CentPv2 were largely reduced and chromosomal distributions were changed. These data indicate centromeric satellite repeats evolve rapidly by changing their copy numbers and chromosomal distributions over a short evolutionary time frame in *Phaseolus* species.
In cowpea, which diverged from common bean ~5 million years ago, we found that the centromeric repeats are not conserved between these two species. DAPI-stained pachytene chromosomes exhibit very large centromeric regions and seven out of 11 centromeres possess a 455-bp tandem repeat, identified using computational approaches. No other highly abundant repeats were found at centromeres. This analysis show centromeric regions have evolved rapidly in both sequences and structure after the divergence of common bean and cowpea and contributes to our understanding of centromere structure and evolution in higher eukaryotes.

W680: Plant Chromosome Biology

Dynamic Histone Phosphorylation Changes for Maize Chromosome Orientation and Segregation

Handong Su[1], Yalin Liu[2], Jing Zhang[2], Qianhua Dong[3], Jing Yuan[4], James A. Birchler[5] and Fangpu Han[6], (1)Institute of Genetics and Developmental Biology, Beijing, China, (2)Institute of Genetics and Developmental Biology, Beijing, China, (3)New York University, New York, NY, (4)Institution of Genetics and Developmental Biology - Chinese Academy of Sciences, Beijing, China, (5)University of Missouri, Columbia, MO, (6)Chinese Academy of Sciences, Beijing, China
Genomic stability is very important for all organisms. It requires a precise cell division with proper orientation and successful segregation of chromosomes in both mitosis and meiosis. Histone modifications can preclude or attract factors that are associated with various functional chromatin organizations. Dynamic histone phosphorylations during cell division are very important for chromosome orientation and segregation. Here, we firstidentified two histone phosphorylation sites at H2AT133 and H3T3 in maize and clonedthe conserved kinases Bub1 and Haspin that catalyze these modifications. Phosphorylation of histone H2A at T133 is specific to centromere regions and associated with centromere activity. H2A phosphorylation is coincident with the signals of kinase Bub1 during meiosis. The H2AT133ph signals reduced in somatic cells of *bub1* RNAi transgenic lines, but there were no changes in meiocytes. ChIP-seq distribution patterns in centromeres were in substantial agreement with anti-CENH3 and anti-H2AT133ph, suggesting that the CENH3 nucleosome has a conserved organization with histone H2AT133 phosphorylation. The signals of Haspin-dependent H3T3 phosphorylation are distributed along the entire sister chromatids in prophase and are restricted to the inner parts of functional centromeres in metaphase. The functional centromere contains a mixture of H3 and CENH3 nucleosomes and changes organization at metaphase. In maize meiocytes, H3T3 phosphorylation extends to the entire chromosome in metaphase I, but it is only observed in the centromere in metaphase II. We also observed Haspin signals along the entire chromosome in pachytene. These results suggest that a series of specific dynamic phosphorylations of core histones within the centromere regions distinguish different centromeric organizations during cell division, which may be vital for successful chromosome segregation.

W681: Plant Cytogenetics

Development of Single Chromosome Genomics in Plants

Petr Capal[1], Nicolas Blavet[1], Pradeep Ruperao[2], Marie Kubalakova[1], Anthony Klein[3], Judith Burstin[4], David Edwards[2] and **Jaroslav Dolezel**[1], (1)Institute of Experimental Botany, Olomouc, Czech Republic, (2)University of Western Australia, Perth, Australia, (3)INRA, Dijon, France, (4)INRA, UMR1347 Agroécologie, Dijon, France

Dissecting nuclear genomes to separate chromosomes can reduce DNA sample complexity, validate genome, assemblies and assess structural genome variation. Flow cytometric sorting is suitable for the purification of large numbers of chromosomes. However, a chromosome of interest must be discriminated from other chromosomes based on karyotype. This is not possible in many species, where some chromosomes can only be sorted into groups of two or more chromosomes. An alternative option is to isolate a single copy of a chromosome and avoid contamination with other chromosomes. With this in mind, we have developed a protocol for the production of microgram quantities of DNA from single copies of flow-sorted chromosomes. Chromosome DNA sequence coverage may be increased by merging sequences obtained from several separate isolations. Our results indicate that single chromosome sequencing is suitable to identify genic sequences on individual chromosomes, develop chromosome-specific DNA markers, verify assignment of DNA sequence contigs to individual pseudomolecules, and validate whole genome assemblies. Moreover, single chromosome sequencing provides opportunities to analyze chromosome structural heterozygosity and determine haplotype phase in plants. By overcoming the need to discriminate individual chromosomes, our approach greatly expands the potential of chromosome genomics, and the method can be applied to any plant species from which chromosome samples suitable for flow cytometry can be prepared.

W682: Plant Cytogenetics

What Role can the Meiotic Axis Play in Stabilising Polyploid Meiosis?

Christopher H. Morgan, University of Birmingham, Birmingham, United Kingdom

Autopolyploid species are widespread within the plant kingdom and offer several putative evolutionary advantages relative to their diploid counterparts. However, in autopolyploids the presence of multiple homologous chromosomes with equal pairing capabilities can result in meiotic irregularities such as multivalent formation and chromosome mis-segregation.

Arabidopsis arenosa has naturally evolved diploid (2n=16) and tetraploid (2n=32) populations and has recently become established as a model organism for investigating autopolyploid meiosis. It has previously been demonstrated that tetraploid *A. arenosa* has reduced crossover rates when compared with the diploid and it is proposed that this enables meiotic stabilisation by inhibiting multivalent formation and promoting bivalent-like pairing. It has also been shown that a number of genes involved in meiotic axis formation, sister chromatid cohesion and homologue synapsis have undergone strong ploidy-specific differentiation in *A. arenosa.* For instance, one protein that has undergone selection in the tetraploid *A. arenosa* lineage is the chromosome axis protein ASY1, the plant functional homologue of *HOP1* in *S. cerevisiae*.

We have used a combination of cytogenetics and immunocytochemistry to further our understanding of how axis formation and synapsis progress in *A. arenosa* and to determine whether any meiotic differences can be observed in tetraploid *A. arenosa* plants expressing either the diploid or tetraploid *ASY1* allele. Using the same techniques we have also examined the effect that reduced crossovers can have on autopolyploid meiosis in *A. thaliana* using colchicine induced neo-tetraploid lines of meiotic axis mutants. Progress in both of these areas will be discussed.

W683: Plant Cytogenetics

Homoeologous Relationships of *Aegilops caudata* and Wheat Chromosomes as Determined by Single Gene FISH Mapping

Tatiana V. Danilova, Bernd Friebe and Bikram S. Gill, Kansas State University, Manhattan, KS

Aegilops caudata L. (=*Ae. markgrafii* (Greuter) Hammer, 2n = 14), a diploid wild relative of bread wheat is a source of agronomically important traits for wheat improvement. *Ae. caudata* is also an interesting example of speciation within the tribe Triticeae because of its highly rearranged genome; its structure and evolution have not been studied in detail. Fluorescence in situ hybridization (FISH) is a useful tool for studying chromosome organization. To overcome the problem of high content of repetitive elements in Triticeae genomes (80-90%), repeats free cDNAs can be used to develop chromosome specific FISH probes. Earlier we had developed a wheat physical map with several cDNA FISH markers located on each of the 14 homoeologous chromosome arms. All wheat cDNA FISH probes produced distinct signals on chromosomes of *Ae. caudata*, and allowed elucidation of their macrostructure and homoeology to wheat. Three submetacentric chromosomes, designated previously as A, C and D are homoeologous and mostly colinear to wheat group 1, 5 and 6 chromosomes respectively; though the long arm of chromosome D has a terminal translocation from chromosome E (group 7). Other chromosomes are highly rearranged: submetacentric B and acrocentric E, F and G contain regions homoeologous to wheat groups 2, 7, 3 and 4 respectively, involved mostly in intrachromosomal rearrangements but some interchromosomal translocations as well. Location of near-centromeric cDNA probes and wheat centromere specific transposable element showed that the centromeres of *Ae. caudata* chromosomes did not change their positions compared with their wheat homoeologs.

W684: Plant Cytogenetics

What Can We Learn about Meiosis by Studying the PCH2 in Maize?

Ljudmilla Timofejeva, Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia

Meiotic mutant *meiN2415* identified in maize by forward genetic approach exhibits defects in chiasma resolution and chromosome segregation at anaphase I (AI). Using Bulk Segregation Analysis, we mapped the mutation to chromosome 10. A point mutation resulting in a premature stop codon was discovered in the *trip13* gene encoding conserved AAA-ATPase family protein THYROID RECEPTOR-INTERACTING PROTEIN13. Expression of mutated *trip13* in *E.coli* revealed truncated protein of ~36 kDa (the wild type PCH2 is ~60 kDa). The maize TRIP13 is a homolog of the mouse TRIP13 and the budding yeast Pch2 (Pachytene checkpoint2). Although TRIP13/Pch2 is conserved in organisms that undergo meiosis, its role is controversial. The maize *pch2* mutant differs significantly from rice and Arabidopsis *pch2* mutants. Transmission electron microscopy revealed incomplete synapsis. Although synapsis is delayed and incomplete, bivalent formation is not impaired in *meiN2415* mutant. Mutation severely affects chromosome segregation at AI resulting in chromosome bridges due to unresolved chiasmata. Surprisingly, immunostaining *meiN2415* meiocytes with antibodies against two structural component of SC lateral elements ASY1 (ortholog of yeast Hop1p) and AFD1 (ortholog of REC8, one of components of cohesion complex) as well as ZYP1, a structural component of the SC central element and a functional ortholog of yeast Zip1p, appeared to be similar to that observed in fertile plants suggesting that maize PCH2 does not regulate the initial chromosomal recruitment of these proteins. Using BiFC technique, we detected transient protein-protein interaction between PCH2 and ZYP1 in tobacco leaves infiltrated with vectors harboring *PCH2* and *ZYP1* coding regions.

W685: Plant Cytogenetics

Meiotic Thermal Tolerance in *Arabisopsis arenosa*

Kevin M. Wright, Harvard University, Cambridge, MA
How does adaptation to external abiotic environments affect meiosis, gamete development, and recombination? Meiosis is a tightly regulated process fundamental to all sexual reproducing organisms. Stabilizing selection upon proper chromosome pairing and crossing-over is strong because the mis-segregation of homologs can result in sterility and/or lethality. Meiotic processes are sensitive to the environment, particularly temperature, in which gametes develop. It is largely unknown whether the meiotic machinery of organisms native to different thermal environments is locally adapted to function at specific temperatures. We investigate this question by measuring pollen fertility and prophase I markers of meiotic progression – synaptonemal complex (SC) development- in populations of *Arabidopsis arenosa* from the hot/dry Pannonian Basin and cool/wet Carpathian Mountains of central Europe. We find that plants from these regions have significantly higher pollen fertility when grown at native versus non-native temperature conditions. The reduction in pollen fertility of plants from the Carpathians at high temperature is associated with the failure of SC establishment. In order to identify candidate genes that may underlie this phenotypic variation, we measured genome-wide genetic differentiation between Pannonian and Carpathians populations. Of the 148 meiosis functioning proteins, only three exhibit significant differentiation between populations. Interestingly all three of them are structural, SC associated proteins. Patterns of genetic variation within populations are consistent with selection in the Pannonian lineage and may contribute to greater thermal tolerance in SC establishment and pollen viability.

W686: Plant Cytogenetics

Three Genome Meiosis in *Brassica* Hybrids

Annaliese S Mason, Plant Breeding Department, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany
The agriculturally important *Brassica* genus contains six species with an interesting genomic relationship: three diploids have genome complements 2n = AA, BB and CC and three allotetraploids have genome complements 2n = AABB, AACC and BBCC. Although no natural allohexaploid *Brassica* with 2n = AABBCC exists, this genome composition is of particular interest for development of a new crop type benefiting from hybrid vigour, as well as for investigation of hybrid speciation processes. However, newly synthesised *Brassica* hybrids containing the A and C genomes usually suffer from poor control of non-homologous chromosome pairing during meiosis, resulting in loss of chromosomes and chromosome fragments and hence failure to establish true-breeding, genomically-stable lines. We investigated meiotic behavior in three-genome interspecific hybrids from crosses between *B. juncea*, *B. napus* and *B. carinata*, including allohexaploid types, using a combination of classical and molecular cytogenetics and high-throughput genotyping approaches. Homologous chromosome pairing was mostly normal despite the different species origins, but non-homologous chromosome interactions (A-B, B-C and A-C) did occur at low frequencies between most chromosomes. Surprisingly, strong selective pressure for 2n = AABBCC euploid karyotypes was not observed, and most chromosomes present in only one copy (univalents) were equally likely to be lost as to be retained in next-generation progeny. Meiotic behaviour was observed to be influenced by both genotype and genomic structure in different allohexaploid genotypes and combinations. Ongoing analysis may detect identification of factors associated with meiotic stability in *Brassica*, and facilitate selection and production of new, stable allohexaploid germplasm.

W687: Plant Dormancy Workshop

Dormancy without Reproduction

Wenqin Wang, Waksman Institute, Rutgers University, Piscataway, NJ, Yongrui Wu, Shanghai Institute of Plant Physiology & Ecology, CAS, Shanghai, China and **Joachim Messing**, Rutgers University, Piscataway, NJ
One of the fastest growing biomass consists of duckweeds. They cover slow moving streams and ponds because of run-offs from fertilizers. These plants grow in nearly every geographic location except in permafrost because they could not survive, when the water freezes. In climates with seasons, they have adapted to an interesting cycle of growth and dormancy. When there is shortage of nutrition in the fall or the temperature drops in the winter, they stop photosynthesis and start accumulating starch, which makes them denser causing them to sink. This switch is accompanied by a morphological change from a leaf-like structure, called frond, to turions. Because the water only freezes on top of the ponds and streams, turions can stay in the unfrozen layer at the bottom. In such a state, turions cease most metabolic activities as seeds do. Similar like seeds, they can switch back to a growth phase in spring, when the temperature rises and the water on top melts. They use starch as an energy source and float to the top, where they resume photosynthesis as fronds. Therefore, the question arises which genes are turned off and on during this cycle and whether they resemble a similar developmental program as terrestrial seed-producing plants have? We used next generation sequencing technology to examine the transcriptome of turion development triggered by exogenous ABA in Greater Duckweed, *Spirodela polyrhiza*, a species, whose genome has been sequenced (Nat. Commun. e3311). Interestingly, there is a resemblance of expression patterns between turions and seeds.

W688: Plant Dormancy Workshop

Extensive Transcriptome Changes during Natural Onset and Release of Vegetative Bud Dormancy in Populus

Haiwei Lu[1], Glenn T. Howe[1], David P. Horvath[2], Palitha Dharmawardhana[3], Henry D. Priest[4], Todd C. Mockler[5] and Steven H. Strauss[1], (1)Oregon State University, Corvallis, OR, (2)USDA-ARS, Fargo, ND, (3)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (4)Donald Danforth Plant Science Center, St. Louis, MO, (5)Donald Danforth Plant Science Center, Saint Louis, MO
Endodormancy is a state of suspended plant growth maintained in harsh environments or even during favorable growing periods. To explore genetic basis of vegetative bud endodormancy, we collected paradormant, endodormant, and ecodormant axillary buds from *Populus trichocarpa* trees growing under field conditions in western Oregon. Of 44,441 gene models analyzed by NimbleGen microarrays, 4% were differentially expressed among at least two dormancy states (false discovery rate p-value < 0.05). Of all differentially expressed genes, 69% were down-regulated from paradormancy to endodormancy, suggesting an overall reduced metabolic activity during endodormancy. Multiple chromatin-associated genes showed different expression levels in three states, and two of them (similar to *SPT*, *SUPPRESSOR OF TY*) were strongly up-regulated during endodormancy. Transcription factor genes that showed atypically increased expression during endodormancy include a gene

that seems to encode a trihelix transcription factor, and genes associated with proteins involved in responses to ethylene, cold, and other abiotic stresses. Analyses of phytohormone-associated genes suggest important changes in responses to ethylene, auxin, and brassinosteroids occur during endodormancy. Weaker or little changes were found in genes associated with salicylic acid, jasmonic acid, gibberellins, abscisic acid, and cytokinin. We identified 315 upstream sequence motifs associated with eight patterns of gene expression, including novel motifs and motifs previously associated with the circadian clock and responses to photoperiod, cold, dehydration, and ABA. Analogies between the regulation of flowering and endodormancy suggest important roles for genes similar to *SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE* (*SPL*), *DORMANCY ASSOCIATED MADS-BOX* (*DAM*), and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*).

W689: Plant Dormancy Workshop

Identification of Candidate miRNAs and their Targets Involved in the Dormancy Regulation of Grapevine.

Shuchi Smita[1], Monica Accerbi[2], Pamela J. Green[2], Senthil Subramanian[1] and Anne Fennell[1], (1)South Dakota State University, Brookings, SD, (2)University of Delaware, Newark, DE

Bud dormancy in grapevines plays an important role for survival under harsh conditions. microRNAs (miRNAs) are small non-coding RNAs molecules (typically 21 or 22 nt) that affect many biological processes by negatively regulating gene expression primarily at the post-transcriptional level. To identify and evaluate the roles of miRNAs in bud dormancy in grapevine, we identified conserved and novel miRNAs using high-throughput sequencing and subsequent bioinformatics analyses. We analyzed small RNA libraries from bud tissue at endodormancy (short photoperiod, SD) or paradormancy (long photoperiod (LD) stages. Results revealed conserved and potentially novel miRNAs, a number of which were differentially expressed between these two dormancy stages. Potential target genes were predicted using an in-house Perl script. We compared RNA-Seq derived transcriptomes of LD and SD bud or shoot tip tissues and identified several differentially expressed genes that could be involved in the dormancy cycle. Inverse expression between miRNAs and the corresponding target genes in paradormancy vs. endodormancy stages indicated potential post-transcriptional regulation of some of these genes by miRNAs. This study has identified several conserved and novel miRNAs with a potential role in the bud dormancy cycle.

W690: Plant Dormancy Workshop

Shifts in Balance Among Phytohormones are the Main Factors Driving Differentiation of Bud to Shoot Growth

Wun S. Chao, Munevver Dogramaci, David P. Horvath, James V. Anderson and Michael E. Foley, USDA-ARS, Fargo, ND

Leafy spurge (*Euphorbia esula* L.) is an herbaceous weed that maintains a perennial growth habit through seasonal production of abundant underground adventitious buds (UABs) on the crown and lateral roots. During the normal growing season, differentiation of bud to shoot growth is inhibited by physiological factors external to the affected structure; a phenomenon referred to as paradormancy. Initiation of shoot growth from paradormant UABs can be accomplished through removal of the aerial shoots (hereafter referred to as paradormancy release). In this study, phytohormone abundance and the transcriptomes of paradormant UABs vs. shoot-induced growth at 6, 24, and 72 hr after paradormancy release were compared based on hormone profiling and RNA-seq analyses. Results indicated that auxin, abscisic acid (ABA), and flavonoid signaling were involved in maintaining paradormancy in UABs of leafy spurge. However, auxin, ABA, and flavonoid levels/signals decreased by 6 hr after paradormancy release, in conjunction with increase in gibberellic acid (GA), cytokinin, jasmonic acid (JA), ethylene, and brassinosteroid levels/signals. Twenty four hr after paradormancy release, auxin and ABA levels/signals increased, in conjunction with increase in GA levels/signals. Major cellular changes were also identified in UABs at 24 hrs, since both principal component and Venn diagram analysis of transcriptomes clearly set the 24 hr shoot-induced growth apart from other time groups. In addition, increase in auxin and ABA levels/signals and the down-regulation of 40 over-represented AraCyc pathways indicated that stress-derived cellular responses may be involved in the activation of stress-induced re-orientation required for initiation of shoot growth. Seventy two hr after paradormancy release, auxin, cytokinin, and GA levels/signals were increased, whereas ABA, JA, and ethylene levels/signals were decreased. Combind results were consistent with different phytohormone signals acting in concert to direct cellular changes involved in bud differentiation and shoot growth; in addition, shifts in balance among these phytohormones at different time points and stress-related cellular responses after paradormancy release appear to be critical factors driving transition of bud to shoot growth.

W691: Plant Dormancy Workshop

Higher Seed Dormancy and ABA Sensitivity Improves Wheat Preharvest Sprouting Tolerance

Shantel A. Martinez[1], Keiko M. Tuttle[1], Kimberly Garland Campbell[2], Arron H. Carter[1] and Camille M. Steber[2], (1)Washington State University, Pullman, WA, (2)USDA-ARS, Pullman, WA

Many white wheat cultivars are susceptible to preharvest sprouting, the germination of mature grain on the mother plant when rain occurs before harvest, due to insufficient seed dormancy. However, cultivars with high seed dormancy have problems with poor germination and emergence if sown within 6 to 8 weeks of harvest. Thus, the breeding goal is to have sufficient seed dormancy at maturity to prevent sprouting, but to lose dormancy rapidly through 6 to 8 weeks of after-ripening during dry storage. Two approaches are being taken to improve preharvest sprouting tolerance without adverse effects on emergence. The first is to deploy *ENHANCED RESPONSE TO ABA8* (*ERA8*), a wheat mutant isolated for increased sensitivity to the dormancy hormone ABA in soft white wheat. The second approach is to perform genome wide association mapping for seed dormancy and seedling emergence within white winter wheat breeding programs. The increased ABA sensitivity in the Zak *ERA8* line is associated with increased seed dormancy compared to wild type soft white spring Zak. However, Zak *ERA8* loses dormancy fairly quickly, within about 8 weeks of after-ripening. Zak *ERA8* showed improved PHS tolerance based on spike wetting tests over 3 field environments. While the effect Zak *ERA8* on germination and field seedling emergence were not statistically significant, future work will examine whether a small decrease in germination capacity impacts grain yield.

W692: Plant Dormancy Workshop

The Role for Jasmonates in Wheat Dormancy Release by Stratification

Jose Barrero[1], Qian Xu[1], Thy T. Truong[2], John V. Jacobsen[1], Charles H. Hocart[2] and Frank Gubler[1], (1)CSIRO Agriculture, Canberra, Australia, (2)Mass Spectrometry Facility, Research School of Biology, Australian National University, Canberra, Australia
Hydration at low temperatures, treatment known as stratification, is widely used to break seed dormancy in many species, including wheat. However, the mechanism by which cold regulates dormancy is largely unknown. Our previous studies showed that jasmonic acid (JA) and its derivative methyl jasmonate promote dormancy release in wheat. We have now found that the stratification of dormant grains is correlated with an increase in endogenous JA content in the embryo. Cold triggers JA production rapidly in dormant embryos after transfer from cold (4°C) to room temperature (20°C). The induction of JA production is dependent on the extent of cold imbibition and is tightly linked with initiation of germination. In addition, we have explored the relationship between JA and abscisic acid (ABA), a well-known dormancy promoter, in cold regulation of dormancy. We found an inverse relationship between JA and ABA content in dormant wheat grains that had been imbibed in the cold and transferred to room temperature. Our results indicate that the action of JA on cold-stratified grains is mediated by suppression of two key ABA biosynthesis genes.

W693: Plant Genome Engineering
TBA
Yinong Yang, Pennsylvania State University, University Park, PA

W694: Plant Genome Engineering
Precision Genome Editing Tools for Non-Transgenic Trait Development
Greg Gocal, Cibus, San Diego, CA
Innovative trait development tools in plant breeding will be crucial for doubling global agricultural productivity by 2050. The *Rapid Trait Development System* (***RTDS***™) is a tool that can deliver precise, non-transgenic and globally acceptable traits. ***RTDS*** employs Gene Repair OligoNucleobases (GRONs) to make defined spelling changes in genomic DNA. We report that ***RTDS*** can significantly improve the outcome of double strand break activity by reliably inducing precise and targeted nucleotide spelling changes closely aligned to the cut site. Our work demonstrates the significance of gene editing to rapidly, precisely and reliably improve crop performance to develop any trait in commercially relevant crop varieties.

W695: Plant Genome Engineering
TBA
Caixia Gao, Institute of Genetics and Developmental Biology, CAS, Beijing, China

W696: Plant Genome Engineering
Editing Crop Genomes to Advance Agriculture
Mark Cigan, DuPont Pioneer, Johnston, IA
Targeted DNA double-strand breaks (DSB) can substantially increase the frequency of genome editing through homology directed repair. In the past decade, several site-directed nucleases have been utilized toward this end, but the recently discovered RNA-guided Cas9 endonuclease has truly revolutionized the field due to its simplicity, activity, and versatility. In plants, the ability to generate targeted DSB has three major applications: gene mutagenesis, gene editing, and site-specific gene insertion. Here we report a successful application of the Cas9-guide RNA system as an efficient tool for all three genome editing approaches in maize. Biolistic transformation of Cas9, guideRNAs, without or with repair templates, faithfully directed sequence alteration at five different genomic regions: upstream of the *liguleless-1* gene, two male fertility genes, MS26 and MS45, and two acetolactate synthase genes, ALS1 and ALS2. Mutations were identified at all target sites often recovering plants with bi-allelic multiplex mutations. DSBs generated Cas9-guide RNA also stimulated gene insertion at a site near *liguleless-1* while gene editing of ALS2 resulted in plants resistant to chlorsulfuron. In addition, transient delivery of guide RNA (as RNA molecules) and repair template directly into immature embryo cells containing pre-integrated Cas9 also resulted in ALS2 editing. These examples represent advances in new breeding technologies and demonstrate the potential of Cas9-guide RNA to accelerate plant biology to help meet demands on global agriculture.

W697: Plant Interactions with Pests and Pathogens
Use of an Autoimmune Mutant to Probe the Maize Defense Response
Peter Balint-Kurti, USDA-ARS, North Carolina State University, Raleigh, NC
The hypersensitive response (HR) is a defense response found in all higher plants in which resistance (R-) proteins in the first infected cells detect the presence of specific molecules associated with the pathogen and initiate a series of defense responses including rapid cell death. In many cases this halts the infection before it can spread to other cells in the plant. Despite its importance in maintaining healthy crops, much remains unknown about the HR. We have used a novel genetic screen involving Rp1-D21, an autoactive R-protein from maize that confers a constitutive HR, to identify loci, genes and pathways associated with natural genetic variation controlling HR in maize. We have shown that proteins encoded by some these of these genes interact specifically with the Rp1-D21 protein to inhibit the HR. Other genes appear to act in a more general way to suppress programed cell death.

W698: Plant Interactions with Pests and Pathogens
Exploring Nonhost Resistance in a Model Legume, *Medicago truncatula* Against Asian Soybean Rust.
Upinder S. Gill, The Samuel Roberts Noble Foundation, Ardmore, OK
Asian Soybean Rust (ASR), caused by *Phakopsora pachyrhizi*, is an economically important disease of soybean and it is considered as a major threat for future food security. In the past, few resistance genes that confer resistance against ASR have been identified and deployed in soybean cultivars but due to heterogeneous population and high diversity among *P. pachyrhizi* races, the host resistance is often short lived. Compared to

host resistance, nonhost resistance is more durable. Nonhost resistance is resistance exhibited by a plant species against a particular pathogen. *Medicago truncatula*, a model legume species, is a nonhost to *P. pachyrhizi* and shows incompatible interactions with this fungus. To better understand nonhost resistance aginst ASR, we performed forward genetics screening of ca. 2,500 *M. truncatula Tnt1* insertion lines and identified several mutants with altered response to ASR. These mutants include phenotypes such as *enhanced penetration of rust* (*epr*), *sporulation of soybean rust* (*spr*) which is normally rare on *M. truncatula*, *enhanced cell death* (*ecd*) upon rust inoculation and *inhibition of rust germ tube* (*irg*). Further characterization of some of these mutants and the mutated gene responsible for the phenotype will be presented.

W699: Plant Interactions with Pests and Pathogens

Different Mechanisms of Whitefly Resistance in Cabbage and its Wild Relatives

Koen Pelgrom[1], Roeland E. Voorrips[1], Colette Broekgaarden[1,2] and **Ben Vosman**[1], (1)Wageningen UR Plant Breeding, Wageningen, Netherlands, (2)Utrecht University, Utrecht, Netherlands

Brassica crops, in particular Brussels sprouts, kale and Savoy cabbage suffer from the cabbage whitefly (*Aleyrodes proletella*). As a first step to identify resistant material we screened a collection of 432 accessions, including wild material and landraces of *Brassica oleracea* as well as wild relatives for whitefly resistance in no-choice field and greenhouse experiments. Resistant accessions were identified among *B. oleracea* var. *capitata* (heading cabbage) landraces and some of the wild relatives, notably *B. incana* and *B. villosa*. While in heading cabbage resistance is only expressed in plants of at least twelve weeks old, some wild relatives were already starting to express resistance at a the age of six weeks. QTL mapping was used to identify chromosomal regions involved in whitefly resistance. Within the heading cabbages, QTLs were found for whitefly adult survival, oviposition rate and morphological traits possibly related to the resistance. However, none of the measured morphological traits co-localized with the whitefly resistance QTLs. Analysis of the probing behaviour of whitefly adults revealed that the resistance trait affects *A. proletella* at the phloem level. Another population that we studied was obtained from a cross between a fully whitefly resistant plant of *B. incana,* which is densely covered with trichomes, and a susceptible *B. oleracea* cultivar. In this population we mapped whitefly resistance to a single locus. At the same locus we also mapped the presence/absence of trichomes. We conclude that several mechanisms for whitefly resistance exist in the Brassica genepool, which may be exploited in breeding.

W700: Plant Interactions with Pests and Pathogens

Integrated Profiling of Histone Modifications and Gene Expression in the Rust Resistance Response of Common Bean.

Vasudevan Ayyappan, Delaware State University, Docer, DE

Vasu Ayyappan, Delaware State University

Histone modifications such as methylation and acetylation play a substantial role in regulating gene expression in normal and stressed plants. On the other hand, transcriptome profiling using RNA-Seq analysis accounts for an array of differentially expressed transcripts between treated and untreated samples. Understanding the epigenomic and transcriptomic regulation of such stress-responsive genes in non-model crops is limited. In this study, we report the genome-wide analysis of histone methylation ($H3K9_{me2}$) and acetylation ($H4K12_{ac}$) in common bean (*Phaseouls vulgaris* L.) under rust (*Uromyces appendiculatus*) stress by using two next-generation sequencing methods - chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA sequencing (RNA-Seq). Using ChIP-Seq analysis, we discovered 1,235 histone methylation and 556 acetylation responsive genes from common bean leaves infected with rust at 0, 12, and 84h. RNA-Seq analysis revealed 145 and 1,763 differentially-expressed genes between mock-inoculated and inoculated plants, respectively. After analysis of the stressed and unstressed transcriptomes, we identified key defense responsive genes and several transcription factors upregulated in stress. The integrated ChIP-Seq and RNA-Seq analysis helped in identification of genes associated with DNA-methylation, histone-methylation, histone-acetylation, chromatin remodeling, and polycomb group of proteins. We classified the transcripts using Gene Ontology (GO) and EuKaryotic Orthologous Groups (KOGs). After assigning the GO terms to pathways, we identified a putative pathway with ten key genes involved in plant-microbe interactions. This is the first comprehensive report on integrated genome-wide profiling of bean-rust interaction and the resources developed here will be beneficial to better understand other epigenomic regulation mechanisms in non-model species under biotic and abiotic stresses.

W701: Plant Interactions with Pests and Pathogens

Grape Leafroll Virus: A Systems Approach to Understand its Interaction with the Plant and its Effect on Fruit Ripening

Laurent Deluc[1], Amanda Vondras[1], Satyanarayana Gouthu[1] and Robert Martin[2], (1)Oregon State University, Corvallis, OR, (2)USDA-ARS, Corvallis, OR

The post-transcriptional regulation of gene expression through miRNA and alternative splicing is an essential component of plant development and response to stress. The present study used RNA and small RNA sequencing technologies to assemble a holistic view of these regulatory agents during ripening in an economically important crop (wine grape) and in response to infection by Grape Leafroll associated Virus (GLRaV). mRNAs and small RNAs from healthy and GLRaV-infected pinot noir fruits were measured weekly from the ripening onset to fruit maturity. We observed approximately 1,700 differentially expressed genes upon comparing healthy and infected berries. Among these genes, a small subset was alternatively spliced (differential exon usage in healthy vs. virus-infected plants). Interestingly, between 1% and 3% of all small RNA reads in GLRaV-infected samples mapped exclusively to the viral genome, providing a trove of information to better understand the interaction between the plant and the virus. Finally, by constructing a network of viral infection-associated differentially expressed genes and incorporating information about changes in the small RNA landscape and the predicted targets of those small RNAs, we identified core nodes through which the presence of the virus potentially signals widespread changes in the berry transcriptome during ripening.

W702: Plant Interactions with Pests and Pathogens

The Role of Light Acclimation and Retrograde Signalling in Regulation of Cell Death and Immune Defences in Higher Plants.

Stanislaw Karpinski, Warsaw University of Life Sciences, Warsaw, Poland

Regulation and optimization of photosynthesis, transpiration, light acclimation and innate immunity is essential for agriculture and forestry plants' productivity. Our results suggest that changes in photosystem II (PS II) maximum efficiency, water use efficiency, hormonal and reactive

oxygen species cellular homeostasis, cell death and seed yield can be defined by the exponential function and simple equation with natural logarithm ($y = y_{0*}e^{-Kx}$), that depends on molecular regulators: *LESION SIMULATING DISEASE 1* (LSD1), *ENHANCED DISEASE SUSCEPTIBILITY 1* (EDS1) and *PHYTOALEXIN DEFICIENT 4* (PAD4). The *LSD1* recessive null mutant (*lsd1*) regardless of permissive laboratory or non-permissive laboratory and field conditions demonstrates constant seed yield, but significant variation in PSII maximum and water use efficiencies, and in foliar transcriptomes that depend on EDS1 and PAD4. Obtained results suggest that LSD1/EDS1/PAD4 constitute at least tree component molecular machinery of chloroplast retrograde signalling regulating plant productivity in the field. Biotechnologically proven potential of these and other cell death regulators in regulation of crops' stress tolerance (drought, UV, pathogen attack, root hypoxia and excess light), biomass quality (cell wall composition and its energy potential) and quantity (trees growth in the field) and seed yield will be presented and discussed. In our recent studies, we have identified a novel transcriptional regulators of the chloroplast retrograde signalling and are assessing the modulation of the redox homeostasis as a new strategy to improve plant performance in stress conditions.

W703: Plant Molecular Breeding

Does marker assisted breeding in common bean have a pulse?

Phillip Miklas, USDA-ARS, Prosser, WA

Common bean (*Phaseolus vulgaris*) representing both dry edible and snap (green or garden beans) bean, is the most widely grown and consumed pulse crop in the world. Beans are produced on 45 M ha worldwide. Cultivar improvement is critical for sustainable production and to meet growing demand for beans. Many bean breeders include some form of marker-assisted breeding in their germplasm and cultivar development pipelines. Marker-assisted selection or detection of specific loci primarily conditioning disease resistance has been commonplace since the early 1990s. Marker-assisted breeding for more complex traits like drought tolerance, quality factors, or yield, has not been developed yet. Marker-assisted breeding for performance traits based on haplotypes across whole genomes is in the exploratory phase. Limited economy of scale may restrict routine application of genomic assisted breeding in bean. A reference genome for common bean was recently published. This new resource has facilitated increased marker densities, GWAS, and candidate gene discovery for simple and complexly inherited traits. Although the number of papers on genomic characterization of traits and QTL have skyrocketed with the advent of these new tools, few validated markers for applied selection of the studied traits have been forthcoming. Many breeders are revisiting old genetic populations with a goal of increasing marker densities (SNPs primarily) for physically mapping important genes and QTL, hopefully leading to markers more tightly linked with their targets and easier to assay for MAS applications. Recent discoveries of markers with application for MAS will be reviewed. Marker-assisted breeding in beans does have a pulse that will continue to get stronger as we learn how to better capitalize on the new genomic tools becoming available.

W704: Plant Molecular Breeding

Mapping-by-Sequencing of Major Genes and QTLs in Allotetraploid Upland Cotton

Gregory Thyssen[1], Marina Naoumkina[2], Hee Jin Kim[2], Md Sariful Islam[2], Doug J Hinchliffe[1], Brian D Condon[1] and David D. Fang[2], (1)Cotton Chemistry and Utilization Unit, USDA-ARS-SRRC, New Orleans, LA, (2)Cotton Fiber Biosciences Unit, USDA-ARS-SRRC, New Orleans, LA

The genomic reference sequences available to the cotton community have grown dramatically over the last few years, tracing an arc that many polyploid species can expect to follow. Draft and reference quality genomes for related diploids were released, followed by draft tetraploid genomes. At each step, we employed new strategies to identify candidate genes for the agronomic traits we study. The creation and sequencing of nearly isogenic lines and bulked segregant populations ensures that genetic diversity is largely limited to the regions under phenotypic selection. However, the presence of similar homeologous sequences in polyploid genomes presents an analytic challenge, as does local rearrangements of chromosomes of the studied cultivars relative to the available reference sequences. When suitable sub-genome read sorting and megabase-scale syntenous reference sequences are available for a genetic locus, simple binning of polymorphisms can identify a diverse region that is closely linked to the gene or QTL. We have converted these polymorphisms to genetic markers that facilitated the fine mapping of genetic loci to the kilobase scale, revealing candidate genes. I will discuss the genetic basis of end-use properties of non-woven fabrics including natural flame resistance. I will present our progress on identification of major cotton fiber property genes and QTLs including short fiber mutants Li_1 and Li_2, the immature fiber mutant *im*, and fiber strength QTLs and look forward to the use of a MAGIC population for allele discovery.

W705: Plant Molecular Breeding

Back to the Roots of Wheat Breeding: How Heading Date Connects to Underground Plant Development

Kai Voss-Fels[1], Matthias Frisch[2], Lunwen Qian[1], Wolfgang Friedt[1], Stefan Kontowski[3], Sven Gottwald[1] and Rod Snowdon[1], (1)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (2)Justus Liebig University, Giessen, Germany, (3)W. von Borries-Eckendorf GmbH & Co. KG, Leopoldshöhe, Germany

An untapped promising potential to increase the stagnating yields of modern wheat and to enhance the productivity for its use as food, feed, and an increasingly important bioenergy resource is believed to lie in the plants' "hidden half", the root system. In wheat, root traits have been associated with higher productivity and yields, improved nutrient acquisition and increased adaptation to abiotic stresses. They furthermore serve as key components in the plants' response towards soil-borne plant pathogens, such as *Fusarium* fungi, that can cause severe yield losses and seed quality reductions. However, the roots have widely been ignored by scientists and breeders, mainly due to the notorious difficulty in obtaining appropriate measurements. Concise knowledge about the genetic background of root development is therefore strongly required, along with reliable diagnostic molecular markers for genomics-based selection of improved, resilient root systems in breeding programs.
We phenotyped an extremely diverse collection of 215 international hexaploid wheat lines in a comprehensive greenhouse screen for basic seedling growth traits. Using genome-wide genotyping data from the 90K Illumina Infinium SNP array, we identified common chromosome regions linking root mass and heading date phenotypes. Our results give interesting insights into relationships between these two traits, discriminating between linkage and pleiotropy, and provide breeders a means to maintain heading characters while maximizing root diversity.

With this respect our findings also provide a valuable basis for genomics-based selection approaches to improve overall resilience in novel high-yielding wheat varieties with extended root systems.

W706: Plant Molecular Breeding

Genome-Wide Association Studies of Disease and Drought-Relative Traits using Whole Genome Re-Sequencing Data in Chickpea

Yongle Li[1], Pradeep Ruperao[2], David Edwards[2], Jacqueline Batley[2], Kristy Hobson[3], Jiayin Pang[2], Kadambot Siddique[2] and Tim Sutton[4], (1)University of Adelaide, Adelaide, Australia, (2)University of Western Australia, Perth, Australia, (3)NSW Department of Primary Industries, Tamworth, Austria, (4)South Australian Research And Development Institute, Adelaide, Australia

Next-generation sequencing (NGS) technology offers a cheap and high-throughput genotyping option to discover genome variation and selection signatures in less utilised crop species, such as chickpea.

We performed whole genome re-sequencing (WGRS) of 64 Australian chickpea varieties, four Indian landraces, and one wild chickpea species (*Cicer reticulatum*) with 5-15X coverage. Alignment of 1.2 billion Illumina paired-end reads to the draft Kabuli genome sequence of chickpea resulted in the identification of over 800,000 SNPs. To handle the high error rate of NGS data, allele frequencies were estimated using site frequency spectrum as prior leading to improved inference of population genetic parameters. Population structure analysis reveals distinct groups among varieties and narrow genetic diversity in recently released varieties. Several regions of the chickpea genome are under positive selection based on Tajima's D test. Both Fst genome scan and genome-wide association studies (GWAS) identify a 100kb region on chickpea chromosome 4 that is significantly associated with ascochyta blight resistance, a fungus disease that severely impacts the chickpea production in Australia and other regions of the world. This region is co-located in a large QTL interval of 7Mb~30Mb confirmed previously by three different mapping populations genotyped at low density with SSR or SNP markers. This 100kb region has been validated by GWAS of another 132 advance lines with ~140,000 SNPs. In total, 13 predicted genes are located in this region including NBS-LRR receptor-like kinase, wall-associated kinase, zinc finger protein and serine/threonine protein kinase. One significant SNP located in the coding sequence of a predicted gene leads to amino acid substitution. This demonstrates the power of combining WGRS data with relatively simple traits in fast developing "functional makers" for marker-assisted selection.

The 132 advance lines were phenotyped with 12 yield and yield relative traits in three drought-prone field environments. GWAS analyses identify 35 SNPs significantly ($p < 3.45e-07$) associated with 6 traits in total. The result of genomic prediction using these lines as training population will be presented as well.

W707: Plant Molecular Breeding

Map-Based Cloning: Novel and Classical Resistance Genes Provide Nonhost Resistance to Wheat Stripe Rust in *Brachypodium distachyon*

Jan Bettgenhaeuser[1], Brian Gilbert[2], Matthew Gardiner[1], Phon Green[1], Magdalena Opanowicz[3], Narayana Upadhyaya[4], John Doonan[5], Michael Ayliffe[2] and Matthew James Moscou[1], (1)The Sainsbury Laboratory, Norwich, United Kingdom, (2)CSIRO, Canberra, Australia, (3)BRC Imaging and Flow Cytometry Facility, London, United Kingdom, (4)CSIRO Agriculture, Canberra, Australia, (5)IBERS, Aberystwyth University, Aberystwyth, United Kingdom

The plant pathogen *Puccinia striiformis* f. sp. *tritici* (PST), commonly known as wheat stripe rust, is an obligate biotroph with a wide host range. In addition to the agronomically important grass species wheat, barley and rye, it can also rarely infect the taxonomically distant *Brachypodium distachyon*. As a model grass species with a sequenced genome, a high rate of recombination, and advantageous morphological characteristics, *B. distachyon* is an excellent system to study the genetic basis of PST resistance in an intermediate nonhost species. We found that most *B. distachyon* accessions were completely resistant to various UK and Australian PST isolates, whereas some accessions showed a range of susceptibility symptoms. Five populations were used to establish the genetic architecture of this resistance, including ABR6 x Bd21, BdTR10H x Tek-4, BdTR13K x Bd21, Luc1 x Jer1, and Foz1 x Luc1. Inoculation with three UK and one Australian PST isolates identified three major QTLs, designated *Yrr1* to *Yrr3*. Interestingly, *Yrr1* was narrowed down to a 75 kb gain of function interval on chromosome 4, which does not contain any known resistance gene homologs. However, classical resistance genes encoding nucleotide binding site, leucine-rich repeat domains (NBS-LRRs) are the candidates underlying *Yrr2* (chromosome 4) and the 102 kb gain of function interval of *Yrr3* (chromosome 2). Strikingly, isolate specificity was observed to *Yrr2*, highlighting the conservation of typical aspects of host-pathogen interaction in this system. These results suggest the involvement of both novel and classical resistance pathways in intermediate nonhost resistance.

W708: Plant Molecular Breeding

Current Status and Future Directions of Molecular Breeding Tools in Rice

Michael J. Thomson, Texas A&M University, College Station, TX

Tremendous advances have been made in molecular genetics and genomics across the rice research community, yet few breeding groups have fully integrated molecular breeding tools into their programs. This talk will review the current status of molecular breeding tools and explore future directions that promise the greatest impact in accelerating progress in plant breeding. Rice is a good example of how basic research has led to practical applications in crop improvement—such as the wealth of cloned genes and QTLs for agronomic traits. The availability of whole genome marker scans using SNP chips and sequencing has also enabled genome-wide association mapping across diverse germplasm panels, leading to valuable information on SNP haplotypes predictive of beneficial alleles for crop improvement. Coupled with low-cost SNP genotyping platforms, these resources can enable more efficient targeted selection at major genes along with genome-wide selection approaches for more complex traits. Even with recent advances, however, major initiatives are still needed to fully realize the potential of molecular breeding tools in mainstream breeding programs. Further efforts are needed in several key areas, including (1) developing resources to share trait-predictive SNP haplotypes and marker validation information for each crop; (2) leveraging advances in high-throughput phenotyping, sequencing, and population development to implement large-scale allele mining of crop genebanks; (3) improving public-sector breeding informatics tools and databases to incorporate high-density SNP data, imputation, and tracking of targeted SNP haplotypes; and (4) exploring the use of genome editing for rapid pyramiding of beneficial alleles at breeding-relevant genetic loci.

W709: Plant Phenotypes

Remote Sensing Technologies for In-Field Specialty Crop Field Phenotyping

Sindhuja Sankaran[1], Jianfeng Zhou[1] and Lav R. Khot[2], (1)Washington State University, Pullman, WA, (2)WSU Center for Precision and Automated Agricultural Systems, Prosser, WA

Unmanned aerial vehicles (UAVs) based high-throughput sensing techniques are being used worldwide for quantification of phenotypic traits in field conditions. UAVs offer versatile sensor integration, rapid and simultaneous data acquisitions with desired high resolution data, and operational flexibility in diverse cropping systems. It also helps in acquiring data even under circumstances where access to field plots using ground platforms can be difficult (e.g. after irrigation/rain event, high biomass). This talk presents some of the on-going field phenotyping research using UAV-based sensing technology and pertinent comparisons to ground-reference data in breeding programs such as legumes and potatoes. Discussion will also focus on the general applications domain and futuristic advancements needed for wider adoption of UAV-based sensing in field phenotyping.

W710: Plant Phenotypes

Phenes and Phenotyping in Maize for Gene Discovery and Breeding

Shawn Kaeppler[1], Nathan Miller[2], Scott C Stelpflug[3], Edgar Spalding[4], Jonathon Lynch[5] and Natalia de Leon[3], (1)Department of Agronomy and DOE Great Lakes Bioenergy Research Center, University of Wisconsin - Madison, Madison, WI, (2)University of Wisconsin, Madison, WI, (3)University of Wisconsin-Madison, Madison, WI, (4)University of Wisconsin Madison, Madison, WI, (5)Penn State University, University Park, PA

Plants respond dynamically to environmental cues throughout growth and development resulting in harvested yield. Novel approaches to plant phenotyping provide information on dynamic responses of plants, and provide insights into novel phenes. Early phenotypes beginning with germination and initial plant growth can be predictive of variation among genotypes at adulthood. Dynamic measurement of kernel imbibition will be provided as an example of a very early phenotype that underlies early seedling growth and thereby potentially stand establishment. Novel phenes provide the potential to identify mechanisms of productivity and stress response. Laser-assisted ablation tomography is a technology for high-throughput analysis of field-grown roots providing new insights into mechanisms of abiotic stress-tolerance in maize. These approaches are examples of possibilities in the emerging area of plant phenotyping. An emerging U.S. initiative, Genomes 2 Fields (genomes2fields.org) will be described. This initiative is a growing, organized effort to make translation of genomic information to productivity in the field the next national priority following the Plant Genome Initiative, supporting the integration of diverse fields including engineering, computation, and biology to lead to predictive models of plant performance in diverse and complex environments.

W711: Plant Phenotypes

A Look Under the Surface: Discovery of Genetic Variation for Root Traits in *Brassica napus* L

Andreas Stahl[1], Derek Wright[1], Wolfgang Friedt[2], Benjamin Wittkop[1] and Rod Snowdon[2], (1)Justus Liebig University, Giessen, Germany, (2)Department of Plant Breeding, Justus Liebig University, Giessen, Germany

Oilseed rape (*Brassica napus* L.) is among the three most important oil crops in the world and plays an integral role in cereal crop rotations. Drought events are among the most important factors limiting yield in oilseed rape production. Especially in light of climate change, adaptation to low rainfall conditions is a major target of ongoing breeding programs. Very little is known about how oilseed rape roots maintain water uptake during rainless growth periods, for example by penetration into deeper soil or increasing the density of fine roots, however these traits are highly relevant to avoid yield losses under dehydration. Moreover, the reaction of the root system to nitrogen availability is also of great interest for increasing nitrogen use efficiency in oilseed rape.

In a large-container phenotyping system, comprising 120 wheelie-bins with 90 cm soil depth, we assessed genetic variation for several root traits in fully mature plants from two diversity sets of 30 genotypes each. The first panel broadly covered genetic variation present in the *B. napus* gene pool, including synthetic accessions and older varieties with high seed erucic acid and glucosinolate content. The second panel represents breeding progress in European commercial elite inbred and hybrid varieties, released to the market between 1989 and 2014. We discovered extreme variation for root traits within the gene pool of *B. napus*, providing unprecedented insight into important diversity for breeding. Besides a differentiation in soil volume penetration and root biomass, contrasting inverse reactions to nitrogen supply were observed in some accessions.

W712: Plant Phenotypes

There’s a World Going on Underground: Imaging Technologies to Understand Root Growth Dynamics and Rhizosphere Interactions

Christopher N Topp, Donald Danforth Plant Science Center, St Louis, MO

W713: Plant Phenotypes

Phenotyping Leaf Biochemical and Physiological Responses to Ozone in Diverse Field-Grown Maize Using Hyperspectral Leaf Reflectance

Craig Yendrek[1], Christopher Montes[2], Tiago Tomaz[3], Gorka Erice[1], Patrick J. Brown[4], Lauren M. McIntyre[5], Andrew D.B Leakey[4] and Elizabeth A. Ainsworth[6], (1)University of Illinois at Urbana-Champaign, Urbana, IL, (2)University of Illinois, Urbana, IL, (3)Visiting Postdoctoral Scholar, Urbana, IL, (4)University of Illinois, Urbana-Champaign, Urbana, IL, (5)University of Florida, Gainesville, FL, (6)USDA-ARS, Urbana, IL

Genetic resources for exploring physiological and agronomic variation in maize are well-established, but there is a field phenotyping bottleneck that limits the rapid and accurate measurement of important traits. Ozone is a damaging air pollutant, known to accelerate leaf senescence and decrease photosynthetic capacity, which ultimately decreases yield. However, measuring biochemical and physiological traits associated with senescence and photosynthesis is slow and laborious. Here, we addressed this challenge by using hyperspectral reflectance spectroscopy (HRS)

to rapidly survey leaf reflectance spectra from over 200 diverse inbred and hybrid maize lines exposed to ambient (~40 ppb) and elevated (~100 ppb) ozone at the Free Air Concentration Enrichment (FACE) field site in Champaign, IL. Partial least squares regression (PLSR) modeling was used to estimate leaf chlorophyll content, percent nitrogen, sucrose content, specific leaf area (SLA) and maximum photosynthetic capacity (V_{max}). PLSR models accurately predicted phenotypes, and there was significant genotypic variation for chlorophyll (15.2-31.2 µg cm^{-2}), nitrogen (3.2-4.3 %), sucrose (3.8-5.7 µmol cm^{-2}), SLA (20.6-29.6 mm^2 mg^{-1}) and V_{max} (18.8-41.1 µmol m^{-2} s^{-1}) within the tested germplasm grown in ambient conditions. Exposure to elevated ozone concentrations decreased chlorophyll by 63.8%, nitrogen by 22.3%, sucrose by 17.5% and V_{max} by 46.8% in the most sensitive genotypes, and increased SLA by 21.2%. There was agreement among years in lines that were responsive to ozone. These results demonstrate the feasibility of using HRS to resolve genotypic variation in maize response to ozone stress, and suggest the approach could be applied to numerous biotic and abiotic stresses.

W714: Plant Phenotypes

Designing Crops for Global Food Security

Andrew G. Sharpe, National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada

The Canada First Research Excellence Fund set up by the Canadian Government recently awarded Cdn$37M to the Global Institute for Food Security (GIFS) at the University of Saskatchewan (U. of S) to establish the Plant Phenotyping and Imaging Research Centre (P^2IRC). The seven years initiative is based on the U. of S campus in Saskatoon and involves researchers at variety of University departments, the Canadian Light Source (synchrotron), National Research Council Canada and Agriculture Canada & Agri-Food Canada. More details about the funding can be found here: http://words.usask.ca/news/2015/07/29/u-of-s-awarded-37-2-million-to-design-crops-for-global-food-security/.

The intent of the funding is to utilize the various expertise on campus to establish a high throughput phenotyping platform in multiple crops with the inclusion of new imaging technologies and strong computational support. Research in the P^2IRC is broken down into fours theme: Theme 1 Phenometrics; Theme 2 Image Acquisition technologies; Theme 3 Computational Informatics of Crop Phenotype Data and Theme 4 Societal and Developing World Impact. A description of the key component of the themes together with details on Theme 1 which comprises high throughput phenotyping, associative genomics and plant pedological phenotypes will be presented.

W715: Plant Reproductive Genomics

Evolutionary Origin of Flowers: A Genome-Wide Study of Floral Development Genes in Seed Plants

Chunce Guo[1,2], Hong Ma[2] and Claude dePamphilis[1], (1)Penn State University, University Park, PA, (2)State Key Lab of Genet Eng, Sch of Life Sci., Fudan Univ., Shanghai, China

Darwin famously characterized the origin and rapid diversification of the angiosperms as an abominable mystery. In recent years, new advances in the fields of molecular phylogenetics and evolutionary developmental biology have not only corrected the mistakes in previous theories and models, but also proposed new and testable hypotheses, revealed possible processes through which the flower first originated, and discussed the contributions of the duplication and diversification of floral developmental genes to the origin of flowers. However, previous studies are insufficient for a complete understanding the contributions of genomic changes to those biological innovations. Here, we examined the origin and duplication history of orthogroups containing 424 genes known to play roles in Arabidopsis floral timing and development, and found that upstream genes in the floral developmental regulatory network had higher frequencies of old orthogroups that originated with tracheophytes, and downstream genes were overrepresented by those in new orthogroups originated within derived angiosperms. Furthermore, the similar pattern was found when using 1312 core-floral genes, which conserved expressed in flower tissues of water lily, rice, California poppy, and Arabidopsis. These results suggest that although much of the floral development genetic toolkit is ancient, gene recruitment has been an ongoing process in floral evolution, and that different components of the floral regulatory network have recruited genes at different rates through history, with a larger proportion of new orthogroups being represented by genes functioning in organ identity, development and maturation than by genes functioning in the environment sensing and floral timing.

W716: Plant Reproductive Genomics

Sex Chromosome Evolution in Haploid Dioecy

Peter Szovenyi[1], Filip Boskovic[1], Adam C. Payton[2] and Stuart McDaniel[2], (1)University of Zurich, Zurich, Switzerland, (2)University of Florida, Gainesville, FL

In diploid dioecious diploid organisms, where sex determination is governed by dimorphic sex chromosomes, sex chromosomes are expected to differ in genetic variability, size, gene content and gene expression due to asymmetric heterozygosity and suppression of recombination. By contrast, in haploid dioecy, sex chromosomes are equally heterozygous and show suppressed recombination in a similar extent. Therefore, if sex-specific evolutionary forces are negligible, sex chromosomes under haploid dioecy are expected to be influenced by similar evolutionary forces and should follow similar evolutionary trajectories. Nevertheless, experimental evidence contradicts this hypothesis in several aspects and suggests that U and V chromosomes differ in size, gene content and level of degeneration. Therefore, we aim at understanding in what extent and why evolutionary trajectories of sex chromosomes under haploid dieocy deviate from theoretical predictions. We investigate this question using the dioicous *Marchantia polymorpha* and its monoicous sister species as a model to gain detailed insights into the evolution of sex chromosomes under haploid dioecy. In particular, I will present results of a genetic cross to identify sex chromosome-linked genes in *M. polymorpha*. Then I will also highlight our comparative genomics effort to understand how sex chromosomes evolved from a putatively monoicous or dioecious ancestor.

W717: Plant Reproductive Genomics

TBA

Meilina Ong Abdullah, Malaysian Palm Oil Board, Kajang, Malaysia

W718: Plant Reproductive Genomics

Abundant Phased, Secondary siRNAs in Plant Reproductive Organs

Blake C. Meyers, University of Delaware, Newark, DE
The major classes of plant small RNAs include microRNAs (miRNAs) and small interfering RNAs (siRNAs); one specific type of the latter are secondary siRNAs, a group well known for including *trans*-acting siRNAs (tasiRNAs). My lab has developed and applied a suite of computational tools and laboratory methods for the analysis of small RNAs and cleaved mRNA targets of small RNAs. By applied these tools to a range of plant species, dissected tissues, and mutants, we have characterized new miRNAs and their targets, and identified diverse populations of phased, secondary siRNAs (phasiRNAs) as well as their miRNA triggers. Our recent work focuses on secondary siRNAs in anthers, particularly grasses in which hundreds or even thousands of long, noncoding RNAs (lncRNAs) are expressed and processed into phasiRNAs. In grass anthers, these lncRNA-derived phased siRNAs are highly enriched, produced by loci dispersed in the genome, and synthesized either pre-meiotically (and are 21 nt) or coordinate with germinal cell maturation for meiosis (and are 24 nt). In addition to their function, our experiments are designed to understand the evolutionary origin and divergence of these phased siRNAs, their lncRNA precursors, and their miRNA "triggers".

W719: Plant Reproductive Genomics

Variation on a Theme: Diploid and Hexaploid Persimmon Regulate Sex in Slightly Different Ways

Isabelle M. Henry[1], Takashi Akagi[2], Ryutaro Tao[2] and Luca Comai[3], (1)University of California, Davis, CA, (2)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (3)Plant Biology and Genome Center, UC Davis, Davis, CA
Plants have evolved many different strategies to prevent self-fertilization and maintain genetic diversity. For example, approximately 5% of plant species are dioecious and bear male and female flowers on separate individuals. One of these species is diploid persimmon, *Diospyros lotus*, a species with heterogametic sex chromosomes similar to humans, with XY male trees and XX female trees. In contrast, commercial persimmons, *D. kaki*, are hexaploid and trees are either fully female or monoeocious, i.e. carrying both male and female flowers. We have determined that sex determination in *D. lotus* relies on a Y-encoded pseudogene, *OGI*, which is expressed in male developing flowers and negatively regulates the expression of an autosomal feminizing determinant called *MeGI*. We will discuss ongoing experiments aimed at deciphering the exact role of *MeGI* in flower development. In *D. lotus*, *MeGI* repression occurs via transitive smRNA from *OGI*. Surprisingly *OGI* expression is elusive in *D. kaki*, but the targeting of *MeGI* by smRNAs is still visible, as is the resulting repression of *MeGI*. Instead, we observed long-term methylation of the *MeGI* promoter and this methylation pattern was fully associated with flower gender. In other words, flowers within a single tree exhibited differential methylation signal at the *MeGI* promoter that was associated with the gender of the flower. This was true even in the case of "sex reversal", cases where a male flower appears on a previously female branch, or vice-versa. In those cases, MeGI promoter methylation was reversed as well, suggesting a role for *de novo* methylation in this process. This suggests that that diploid and hexaploid persimmons are using a slightly different approach to regulating sex determination.

W720: Plant Science at the JGI and KBase

Joint Genome Institute Plant Science Program

Jeremy Schmutz[1], Kerrie W. Barry[1], David M. Goodstein[1], Jane Grimwood[2], Uffe Hellsten[1], Jerry Jenkins[3], John P. Vogel[4], Gerald A. Tuskan[5] and Daniel S. Rokhsar[1], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)DOE Joint Genome Institute, Huntsville, AL, (4)Department of Energy Joint Genome Institute, Walnut Creek, CA, (5)Oak Ridge National Laboratory, Oak Ridge, TN
The Department of Energy Joint Genome Institute is funded to enable scientific advances that benefit the DOE research areas of bioenergy, global carbon cycling and biogeochemistry. These are accomplished through collaborative projects with JGI users through the Community Sequencing Program and the three funded DOE BioEnergy Research Centers. The Plant Program is part of the JGI dedicated to applying advances in genomic technologies for understanding fundamental plant biology through comparative genomics and targeted experiments. Our major goal, in collaboration with plant scientists, is to apply this understanding from genomics to accelerate the improvement and domestication of biofuel crops. The JGI Plant program has produced many of the high-quality reference plant genomes available today and we continue to curate and make available comparative data and analysis via www.phytozome.net. Recently, the plant program has focused on projects that elucidate function of genes through comparative transcriptomics and directed experiments in our JGI Plant Flagship genomes. For these Plant Flagship genomes, we are continuing to improve the accuracy and completeness of the genome sequence and add data to update and improve the reference annotation. We continue to sequence de novo genomes as comparators to the Plant Flagships, have introduced new advances into these pipelines, and have expanded our efforts on projects that use diversity of natural or structured populations to identify and link genotypes to phenotypes for plant traits important in biofuel crops.

W721: Plant Science at the JGI and KBase

The DOE Systems Biology Knowledgebase: Introduction to KBase for Plant Researchers

Robert W. Cottingham, Oak Ridge National Laboratory, Oak Ridge, TN
The U.S. Department of Energy Systems Biology Knowledgebase (KBase, http://kbase.us) provides a computational environment to meet the key challenges of systems biology: predicting and ultimately designing biological function. KBase is distinguished as a knowledgebase that supports the sharing and integration of biological data and any related analysis, modeling, and simulation, not simply a database or a workbench that serves data with canned analyses.

This introduction will briefly cover plant resources available in KBase. KBase does integrate commonly used core tools along with reference and experimental data, and overlays them with new capabilities for visualization, exploration, and predictive analysis designed to accelerate understanding of plants, microbes, and their communities. It provides open access to quality-controlled data and high-performance modeling and simulation tools that enable researchers to build new knowledge, interpret missing information necessary for predictive modeling, test hypotheses, design experiments, and share findings that can be reproduced and extended by others.

User-furnished data can be uploaded, analyzed using high-performance bioinformatics tools, and overlaid visually and analytically on KBase-provided data.

KBase online tutorials such as Build Plant Metabolic Model found at http://kbase.us/tutorials/ provide a good starting place to understand how KBase might be useful in your research and how to get started.

W722: Plant Science at the JGI and KBase

JGI Plant Gene Atlas: Adding Experimentally Derived Functional Annotations to JGI Plants

Avinash Sreedasyam[1], Christopher Plott[1], Morgan Qualls[1], Jerry Jenkins[1], Jane Grimwood[1], Joseph W. Carlson[2], David M. Goodstein[2], Thomas Juenger[3], Yuhong Tang[4], Gerald A. Tuskan[5], Thomas P. Brutnell[6], Sabeeha S. Merchant[7], Stefan A. Rensing[8], John Mullet[9], Todd C. Mockler[10], Gary Stacey[11] and Jeremy Schmutz[1,2], (1)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)University of Texas, Austin, TX, (4)The Samuel Roberts Noble Foundation, Ardmore, OK, (5)Oak Ridge National Laboratory, Oak Ridge, TN, (6)Enterprise Institute for Renewable Fuels Donald Danforth Plant Science Center, St. Louis, MO, (7)University of California, Los Angeles, CA, (8)Faculty of Biology, University of Marburg, Marburg, Germany, (9)Texas A&M University, College Station, TX, (10)Donald Danforth Plant Science Center, Saint Louis, MO, (11)University of Missouri, Columbia, MO

The JGI Plant Gene Atlas is a large, coordinated effort to add functional information to JGI Plant Flagship Genomes. The primary goal of this project is to develop dense RNA-seq data sets for Plant Flagships to form reference transcriptomes across common tissues and conditions and to provide the ability to compare expression across conditions within a plant and from orthologous genes across the JGI Plant genomes. This substantial amount of updateable transcriptomic resource is directly available to JGI users through the JGI Plant Portal at phytozome.jgi.doe.gov. The secondary goal is to provide a technology test bed to further develop genomic techniques that illuminate the function of plant genes and plant regulatory and pathway elements. Gene Atlas project currently comprises 1,078 samples from twelve JGI Plant Flagship genomes: *Chlamydomonas* (algal model), *Physcomitrella* (moss model), *Brachypodium* (a C3 grass model), switchgrass (a woody perennial crop plant), Hall's panicgrass (grass model), *Setaria italica* (grain and forage crop), *Setaria viridis* (model C4 grass), *Sorghum* (a C4 grass bioenergy crop and model) [6 monocots], *Arabidopsis* (model for plant genetics and biology), soybean (legume model and crop plant), *Medicago* (legume model) and poplar (biomass tree crop).

We will present results from our genome version controlled data analysis pipelines which facilitates identifying gene clusters showing tissue/condition specific expression patterns, visualize expression profiles of gene sets across Gene Atlas plants and adding experimental and orthologous based functional annotations.

In addition to standard tissues and conditions, the Gene Atlas includes a comparative condition across the plants focused on nitrogen metabolism providing a single condition that can be compared broadly from the minimal *Chlamydomonas*, into the early plant model *Physcomitrella*, and through the dicots and monocots included in the study. Our results on transcriptional modulation in response to changes in available nitrogen source and conserved gene expression networks across the phylogeny will also be discussed.

W723: Plant Science at the JGI and KBase

Progress Towards an Engineering Quality Sorghum Reference Genome Sequence

John Mullet[1], Jeremy Schmutz[2,3], Jerry Jenkins[2], David Sims[2], Jane Grimwood[2], Brian McKinley[1], Daryl Morishige[1], Brock D. Weers[1], Ashley Mattison[1], Ryan McCormick[1] and Sandra Truong[1], (1)Texas A&M University, College Station, TX, (2)HudsonAlpha Institute for Biotechnology, Huntsville, AL, (3)DOE Joint Genome Institute, Walnut Creek, CA

Sorghum bicolor is a genetic model system for the design of C4 grass crops. Sorghum hybrids are drought resilient annual crops used for production of grain, forage, sugar and biomass. A high quality genome reference sequence is required for gene and regulatory element annotation, accurate sequence variant profiling, ENCODE analysis, and genome engineering. In this project, the quality of the sorghum reference sequence was upgraded using advanced sequencing technology, genome sequence order and completeness was increased using a high resolution genetic map, and genome annotation improved using information from deeply sequenced RNAseq libraries. The improved sorghum V3 reference genome sequence release includes 351Mb of high quality finished sorghum gene space sequence that was generated using manual review, directed primer walk sequencing for gap closure, transposon annotation for tandem repeat resolution, and targeted whole BAC/fosmid based sequencing. Sequence order in the final assembly was improved using a high-resolution genetic map constructed from 10,789 Digital Genotyping GBS-markers ordered using breakpoints present in 400 RILs derived from a cross of BTx623 (reference) and IS3620C. Previously unassembled super contigs were integrated into the reference genome sequence guided by genetic analysis. Overall reference sequence quality was further improved by whole genome sequencing on Illumina platforms. The genome was annotated using deep RNAseq data derived from roots, leaves, stems, and panicles collected at different stages of development. The resulting sorghum RNA Atlas is being used to identify genes expressed in stems that are involved in growth, cell wall biology, and the accumulation of non-structural carbohydrates such as sucrose.

W724: Plant Science at the JGI and KBase

The Pan-Genome of *Brachypodium distachyon*, Capturing the Full Genetic Complement of a Plant Species

John P. Vogel[1], Sean Gordon[1], Bruno Contreras-Moreira[2,3], David L. Des Marais[4], Diane Burgess[5], Wendy Schackwitz[1], Ludmila Tyler[6], Joel Martin[1], Daniel Woods[7], Richard Amasino[7], Anna Lipzen[1], Shengqiang Shu[1], Jeremy L. Phillips[1], Kerrie W. Barry[1], Ana Caicedo[8], Luis Mur[9], Michael R Freeling[5] and Pilar Catalan[10], (1)DOE Joint Genome Institute, Walnut Creek, CA, (2)Estación Experimental de Aula Dei (EEAD-CSIC), Zaragoza, Spain, (3)Fundación ARAID, Zaragoza, Spain, (4)Harvard University, Boston, MA, (5)University of California, Berkeley, CA, (6)Biochemistry and Molecular Biology Department, Amherst, MA, (7)University of Wisconsin-Madison, Madison, WI, (8)Biology Department, University of Massachusetts, Amherst, MA, (9)Aberystwyth University, Aberystwyth,, Wales, (10)Department of Agriculture (Botany), High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain

The genetic diversity of a species is the sum of the diversity found in all individuals of that species. Many studies have attempted to estimate the diversity of plant species by resequencing diverse accessions and aligning the reads to a reference genome. While this approach readily identifies SNPs and small indels with respect to the reference genome, it underestimates total genomic diversity because highly divergent regions align

poorly to the reference and, of course, any sequence not found in the reference will be missed entirely. Thus, the true extent of diversity within plant species is largely unknown. De-novo genome assemblies and annotation can be used to more accurately estimate the true genomic diversity within a species. We applied this approach using 54 *Brachypodium distachyon* accessions to create a pan-genome that contains all the diversity found in the accessions sequenced. Analysis of this data yields a high-confidence *B. distachyon* pan-genome that includes 13,408 core gene clusters found in all lines, 7,283 soft-core genes cluster absent from a few lines, and 17,195 shell gene clusters found in 3 to 52 lines. We find 30,691 gene clusters represented by the reference genome/reference control and a varying number of other genomes. In addition, we identify 7,135 gene clusters not represented in the reference line or controls but present in multiple divergent lines. We show that non-core genes are expressed at lower levels, have narrower and more variable expression across accessions, are evolving faster, have reduced orthology to related grasses and are less likely to have a homeologous gene retained from the ancient genome duplication in the grass lineage. We evaluate the relationship between the number of sequenced lines and their phylogenetic position in relation to the addition of both genic and non-coding sequence. We describe the physical chromosome position of non-core and non-reference genes and its relation to transposable elements. This analysis suggests possible mechanisms by which dispensable genes are eliminated and also barriers to their removal.

W725: Plant Science at the JGI and KBase

Adaptation Proceeds *via* Selection on Pre-Existing Genetic Variation in *Mimulus guttatus*

Kevin M. Wright, Harvard University, Cambridge, MA

Do independent populations evolve to response to similar selective pressures via convergent genetic mechanisms? If so, is convergence due to independent mutations or shared ancestral variation? We address these questions investigating a classic example of adaptation - the colonization of plant species to heavy-metal contaminated soils. We use field-based reciprocal transplant experiments to demonstrate that mine alleles at a major copper tolerance QTL, *Tol1*, are strongly selected in the mine environment, but are neutral, or nearly so, in the off-mine environment. We assemble the genome of a mine adapted genotype and identify regions of this genome in tight genetic linkage to *Tol1*. We discover *Tol1* candidate genes that exhibit significantly large differences in expression between tolerant and non-tolerant genotypes or in allele frequency between mine/off-mine population pairs. We identify a single gene, a multicopper oxidase, which exhibits *both* large differences in expression and allele frequency. Patterns of genetic variation at the five loci with the greatest difference in allele frequency between populations, including the multicopper oxidase, are consistent with selection acting upon beneficial haplotypes that predates the existence of the copper mine habitat. We estimate the age of selected *Tol1* haplotype to be at least 1700 years old and was at a frequency of 0.4-0.6% in the ancestral population when mining was initiated 150 years ago. These results suggest that adaptation to the mine habitat routinely occurs via selection on ancestral variation, rather than independent *de-novo* mutations or migration between populations.

W726: Plant Science at the JGI and KBase

When Biologists and Modelers Meet in KBase: Case Example of Modeling Plant-Microbe Metabolic Interactions

Dave Weston, Oak Ridge National Laboratory, Oak Ridge, TN and **Samuel M. D. Seaver**, Argonne National Laboratory, Argonne, IL

Peatland ecosystems store vast amounts of terrestrial carbon as dead organic peat. The moss plant Sphagnum is a keystone genus in these nitrogen limiting peatland systems, highlighted by an important association with N2-fixing diazotrophic bacterial associates that may ultimately influence Nitrogen (N) and Carbon (C) cycling. However, the reports on the role of diazotroph N-fixation in supporting plant growth and maintaining the ecosystems as C-sinks rather than C-sources are inconsistent.

A modeling framework is needed to explore and understand this uncertainty surrounding N-fixation and N-cycling in such diazotrophic-plant interactions. Here we present one of the first full-scale stoichiometric models describing a simple metabolic interaction between the recently sequenced *Sphagnum fallax* and *Anabaena spp.*, a representative cyanobacterium. We utilized the DOE Systems Biology Knowledgebase (kbase.us) to bring together experimental and computational results in developing the model. The data, the methods to analyze it, and the results are all accessible in a publicly available Narrative on KBase. We demonstrate how our stoichiometric model enabled us to explore the uncertainty surrounding the contribution of fixed N on Sphagnum biomass by formulating strong model predictions that can be tested experimentally.

W727: Plant Transgene Genetics

Improving Economic Traits in Cultivated Potato

David Douches, Michigan State University, East Lansing, MI

Conventionaly bred potato varieties have weaknesses in production or storage that are minimized through management practices or chemical inputs. Numerous genes have been identified that have potential to improve current potato varieties. We are exploring the overexpression or silencing of genes to improve potato varieties that will improve the economics of managing pathogens in the field, improve storage management and improve market quality traits. We have overexpressed eIF4E from *Solanum habrochaites* in PVY susceptible potato varieties. We were able to acheive resistance to three PVY isolates (O, N:O and NTN) in inoculated greehouse tests and an aphid-vectored field study (N:O). We expressed *Rpi-blb1* in conventionally-bred late blight resistant lines as a strategy to enhance resistance to *Phytophthora infestans.* In lab and greenhouse tests we observed resistance across four different isolates of *P. infestans* (US8, US22, US23 and US24) and US23 in inoculated field studies. We reduced expression of the vacuolar acid invertase gene to reduce the conversion of sucrose to reducing sugars in potato tubers. We were able to identify events that allowed the potato tubers to be stored at 4C and maintain industry acceptable chip-processing color. These genes and others offer commercial opportunities to release and breed improved potato varieties that can save costs through production and management practices, hence reducing the environmental impact of growing and storing potatoes.

W728: Plant Transgene Genetics

A Suite of Crop Promoters with Precise Organ-Specific Expression Patterns

Roger Thilmony, USDA-ARS, Albany, CA

The successful deployment of crop biotechnology to improve disease resistance, abiotic stress tolerance, or other desirable traits requires the ability to precisely control the expression of transgenes. Currently, only a few transgene expression control elements (i.e. promoters), which drive gene expression in a tissue- or organ-specific manner, are available from crop plants. We have identified and characterized a collection of promoters that confer distinct leaf-, root-, pollen-, seed- or fruit-specific patterns of expression in transgenic plants. The candidate promoters were selected based on gene expression profiling data and the corresponding upstream promoter sequences were fused to a reporter gene and transformed into rice, *Brachypodium*, wheat, tobacco, tomato and/or Arabidopsis plants. Two green tissue specific promoters predominantly expressed in rice leaves show light responsiveness and four additional rice promoters exhibit cell type specificity within the roots of transgenic plants. Three rice promoters with pollen-specific expression have been characterized and used to successfully ablate transgenic pollen in *Brachypodium*. In addition, five fruit-specific promoter candidates from citrus have been isolated and characterized. Together, these novel promoters form a useful molecular tool box enabling precise, spatially-defined transgene expression of introduced traits within biotech crops.

W729: Plant Transgene Genetics

Transgenic Approaches to Powdery Mildew Resistance in Grapevine

Lance Cadle-Davidson, USDA-ARS Grape Genetics Research Unit, Geneva, NY

W730: Plant Transgene Genetics

A CRISPR-Library Approach for Targeted Knockout of the LRR-RLK XII Gene Family in Tomato

Thomas Jacobs, Patricia Keen, Noe Fernandez-Pozo, Joyce Van Eck and Gregory Martin, Boyce Thompson Institute for Plant Research, Ithaca, NY

Large collections of knockout mutant lines are broadly useful for functional genomics studies. However, in tomato, there is no such large-scale collection. The CRISPR system is a powerful and efficient tool for inducing targeted DNA deletions in plant genomes. The ease with which new targeting vectors can be generated, and the overall efficiency of mutagenesis, enables high-throughput plant transformation schemes. Here, we report the development of a transformation method using a library of CRISPR vectors to efficiently generate a collection of knockout mutants. For proof-of-concept, a CRISPR library was constructed to target the 55-member LRR-RLK XII gene family in tomato. Three gRNAs were designed for each gene, for a total of 165 individual targeting vectors. The targets were synthesized and pooled into a single cloning reaction to generate the library. The library was transformed *en masse* into tomato cotyledons. The incorporated targets were identified by Sanger sequencing. Typically, each transgenic event contains a single, unique targeting vector. Genome modifications were confirmed in approximately 66% of the T_0 events. From a single transformation experiment, ~15 genes from the LRR-RLK XII gene family were modified. This efficient targeted-mutagenesis approach will enable the generation of large collections of mutant tomato lines.

W731: Plant Transgene Genetics

Expanding the Utility and Host Range of *Ensifer*-Mediated Transformation (EMT): A Novel Platform for Engineering Plant Genomes

Manuel A. Lopez Vernaza, Teagasc, Carlow, Ireland

To identify new plant associated bacteria strains that could be used as a substitute for *Agrobacterium tumefaciens* in existing transforming protocols, a functional screen using the open source vector pCAMBIA5105 was performed. “OV14”, a strain of the Gram negative, non pathogenic species *Ensifer adhaerens* was successfully isolated and has been confirmed to have the ability to stably integrate T DNA into dicotyledonous and monocotyledonous genomes[1, 2]. Comparative sequencing of the 7.7Mb genome of OV14 (versus *A. tumefaciens* C58) revealed that the OV14 genome contains two plasmids and two chromosomes in which several homologs to chromosomally based *Agrobacterium* genes that support T-DNA transfer were present[3]. Antibiotic profiling of OV14 showed its susceptibility/resistance to 14 antibiotics, with a strong resistance to kanamycin identified[4]. This is currently being addressed by editing (via CRISPR/Cas9) two homologs (AHK42288 and AHK42618) whose transcription was significantly elevated within 2h exposure to kanamycin[4]. To date the confirmed host range of *Ensifer*-Mediated Transformation (EMT) includes potato, tobacco, canola, rice, cassava and safflower (verified via southern/sequencing) with transient assays indicating potential with wheat, barley, maize and soybean. Present research is focussed on (i) expanding the host range of EMT (ii) confirming EMT as a suitable platform for crop genome editing tools and (iii) examining the presence/absence of genotype dependency in select crops.

1: Wendt et al., *Transgenic research* 21, no. 3 (2012): 567-578
2: Zuniga-Soto et al., *SpringerPlus* 4.1 (2015): 1-10
3: Rudder et al. BMC Genomics, (2014): 15, 268, 1-17.
4: Rathore et al., *FEMS Microbiology Letters* 362.17 (2015): fnv126.

W732: Polyploidy

Multiple Whole Genome Duplications during the Evolution of Hexapods

Michael S. Barker, University of Arizona, Tucson, AZ

W733: Polyploidy

A Polyploid Origin for Dopamine Receptors Across the Vertebrates

Asher K Haug-Baltzell[1], Erich Jarvis[2], Fiona McCarthy[1] and Eric Lyons[1], (1)University of Arizona, Tucson, AZ, (2)Duke University Medical Center, Durham, NC

Dopamine is an important central nervous system transmitter that functions through two classes of receptors (D1 and D2) to influence a diverse range of biological processes in vertebrates. With roles in regulating neural activity, behavior, and gene expression, there has been great interest in understanding the function and evolution dopamine and its receptors. In this study, we use a combination of sequence analyses, microsynteny analyses, and phylogenetic relationships to identify and characterize both the D1 (DRD1A, DRD1B, DRD1C, and DRD1E) and D2 (DRD2, DRD3, and DRD4) dopamine receptor gene families in 43 recently sequenced bird genomes representing the major ordinal lineages across the

avian family tree. We show that the common ancestor of all birds possessed at least seven D1 and D2 receptors, followed by subsequent independent losses in some lineages of modern birds. Through comparisons with other vertebrate and invertebrate species we show that two of the D1 receptors, DRD1A and DRD1B, and two of the D2 receptors, DRD2 and DRD3, originated from a whole genome duplication event early in the vertebrate lineage, providing the first conclusive evidence of the origin of these highly conserved receptors. Our findings provide insight into the evolutionary development of an important modulatory component of the central nervous system in vertebrates, and will help further unravel the complex evolutionary and functional relationships among dopamine receptors.

W734: Polyploidy

Convergent Evolution and Allopolyploid Speciation in the Family of the Model Moss, Physcomitrella

Stefan A. Rensing, Faculty of Biology, University of Marburg, Marburg, Germany

The *Physcomitrium-Physcomitrella* species complex harbours many lineages of polyploid origin and nature. Phylogenetic analyses of nuclear and SSR markers confirm a polyphyletic origin for three cryptic *Physcomitrella* species. Differences in the conservation of mitochondrial editing sites further support hybridization and cryptic speciation within this species complex. Secondary reduction of sporophyte complexity in these species is probably due to the establishment of an ecological niche, namely spores resting in mud and spore dispersal by migratory birds. Besides the *Physcomitrium-Physcomitrella* species complex, the Funariaceae are host to their type species, *Funaria hygrometrica*, featuring a complex sporophyte morphology. We shall present novel data on trait evolution (sporophyte reduction) within the Funariaceae.

W735: Polyploidy

Origin of Polyploidy in Sequoia: Tous pour un, pas de deux, or ménage à trois?

Alison Dawn Scott, University of Wisconsin, Madison, Madison, WI

Coast redwoods (*Sequoia sempervirens*) are well-known for their great height (over 100m) and advanced age (over 2,000 years), but less so for being the only hexaploid conifer (2n=6x=66). Though these colossal ancient trees are limited to the foggy coastal forests of central and northern California and southwestern Oregon, the redwood fossil record suggests a broader historical range across the Northern hemisphere. How, when, and where polyploidization took place remains a mystery, though diverse putative genome donors and polyploidization mechanisms have been proposed. We estimated phylogenetic trees to compare relationships among gene copies in *S. sempervirens* and its close relatives, in order to reveal whether hybridization with an extant lineage contributed to polyploidy in the *Sequoia* lineage. Our results indicate *Sequoia* is an autopolyploid, or that allopolyploidy occurred within the California redwood clade. Our results also suggest duplicate gene pairs in *Sequoia* are much younger than expected based on fossil estimates of whole genome duplication, which may be attributed to continued multisomic inheritance in *Sequoia*. To address this, we consider the role of diploidization in polyploid evolution, and whether slow diplodization helps explain why polyploid gymnosperms are so rare.

W736: Polyploidy

The Double Reduction Landscape in Tetraploid Potato and its Implications for the Genetic Analysis of Autotetraploids

Peter M Bourke, Roeland E. Voorrips, Richard Visser and Chris Maliepaard, Wageningen UR Plant Breeding, Wageningen, Netherlands

The creation of genetic linkage maps in polyploid species has been a long-standing problem for which various approaches have been proposed. In the case of autopolyploids, a commonly-used simplification is that random bivalents form during meiosis. This leads to relatively straightforward estimation of recombination frequencies using maximum likelihood from which a genetic map can be derived. However, autopolyploids such as tetraploid potato (*Solanum tuberosum* L.) may exhibit additional features such as double reduction, not normally encountered in diploid or allopolyploid species. In this study we produced a high-density linkage map of tetraploid potato and used it to identify regions of double reduction in a bi-parental mapping population. The frequency of multivalents required to produce this degree of double reduction was determined through simulation. We also determined the effect that multivalents or preferential pairing between homologous chromosomes have on linkage mapping. Low levels of multivalents or preferential pairing do not adversely affect map construction when highly-informative marker types and phases are used. We reveal the double reduction landscape in tetraploid potato, clearly showing that this phenomenon increases with distance from the centromeres.

W737: Polyploidy

Deciphering the Post-Neotythic Oilseed Rape Genome Reveals the Fascinating Diversifying Force of Polyploidy

Boulos Chalhoub, URGV-INRA, Evry, France

W738: Population and Conservation Genomics

Evolutionary Genomics of Introduced Salmonid Species in Patagonia, South America

Shawn Narum, Columbia River Inter-Tribal Fish Commission, Hagerman, ID, Pablo Gallardo, Universidad de Magallanes, Punta Arenas, Chile, Cristian Correa, Universidad Austral de Chile, Valdivia, Chile, Amanda Matala, CRITFC, Hagerman, ID and Louis Bernatchez, Université Laval, Québec, QC, Canada

Invasive species have become widespread in aquatic environments throughout the world, yet there are few studies that have examined the genetic mechanisms that allow for successful invasion of newly colonized environments. In this study we contrast genomic variation in two salmonid species (*Oncorhynchus tshawytscha*, 11,579 SNPs; *Salvelinus fontinalis*, 13,522 SNPs) with differing invasion success after introduction to new environments in South America relative to populations from their native range in North America. Estimates of genetic diversity were not significantly different between introduced and source populations for either species, indicative of propagule pressure that has previously been shown to maintain diversity in founding populations of invasive species relative to their native range. Introduced populations also demonstrated higher connectivity and gene flow than those in their native range. Evidence for local adaptation was observed within specific introduced populations, and appeared to be driven largely by selection on standing genetic variation. Overall, patterns of genomic variation were consistent with broad or narrow life history variation in each species, and in combination, these factors contribute to invasion success of these two

salmonids to novel environments. These results provide further understanding of adaptation and invasion success of introduced species and may help guide management strategies.

W739: Population and Conservation Genomics

Linking Genes and the Environment: Differential DNA Methylation in Regulatory Regions is Associated With Rainbow Trout Migratory-Related Divergence

Melinda Baerwald, University of California, Davis, Davis, CA

Migration is essential for the reproduction and survival of many animals, yet little is understood about its underlying molecular mechanisms. We used the salmonid *Oncorhynchus mykiss* to gain mechanistic insight into smoltification, which is a morphological, physiological, and behavioral transition undertaken by some juveniles that culminates in a seaward migration. Given that gene x environment interactions and phenotypic plasticity are likely critical components to migratory divergence, epigenetic modifications (e.g., DNA methylation) may regulate gene expression and partially underlie migratory phenotypic diversity. To explore this, we quantitatively measured genome-scale DNA methylation using reduced representation bisulfite sequencing of F_2 siblings produced from a cross between steelhead (migratory) and rainbow trout (non-migratory) lines. We identified 57 differentially methylated regions (DMRs) between steelhead and rainbow trout juveniles. DMRs were of high magnitude, ranging from 20-62% differential methylation between life history types, and over half of the gene-associated DMRs were in transcriptional regulatory regions. Many of the DMRs encode proteins with activity relevant to migratory-related transitions (e.g. circadian rhythm pathway, nervous system development, protein kinase activity). This study indicates that differential DNA methylation at gene regulatory elements may be a critical molecular mechanism allowing interactions between an organism and its environment to balance phenotypic stability and plasticity, and ultimately modify migratory-related phenotypes.

W740: Population and Conservation Genomics

Signatures of Selection Among de novo Assembled Transcriptomes of Four White Pine Species

U. Uzay Sezen[1], Ethan A.G. Baker[1], Taylor Falk[1], Patricia E. Maloney[2], Detlev R. Vogler[3], Annette Delfino-Mix[4], Camille E. Jensen[2], Jeffry B. Mitton[5], Jessica Wright[6], Brian Knaus[7], Hardeep S. Rai[8], Richard Cronn[9], Daniel Gonzalez-Ibeas[1], Hans Vasquez-Gross[10], Randi Famula[11], Jun-Jun Liu[12], Lara M. Kueppers[13], Steven L. Salzberg[14], Charles H. Langley[15], Kristian Stevens[15], Daniela Puiu[16], Aleksey Zimin[17], James A Yorke[17], Marc Crepeau[15], Carol Loopstra[18], Jeffrey Puryear[19], David Neale[11] and Jill Wegrzyn[1], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)University of California, Davis, CA, (3)USDA Forest Service, Placerville, CA, (4) Institute of Forest Genetics, USDA Forest Service, Placerville, CA, (5)University of Colorado, Boulder, CO, (6)USDA Forest Service Pacific Southwest Research Station, Davis, CA, (7)USDA Agricultural Research Service, Corvallis, OR, (8)Utah State University, Logan, UT, (9)USDA Forest Service Pacific Northwest Research Station, Corvallis, OR, (10)University of California Davis, Davis, CA, (11)Dept. Plant Sciences University of California Davis, Davis, CA, (12)Canadian Forest Service, Victoria, BC, Canada, (13)University of California, Merced, CA, (14)Johns Hopkins School of Medicine, Baltimore, MD, (15)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (16)Johns Hopkins University, School of Medicine, Baltimore, MD, (17)University of Maryland, College Park, MD, (18)Texas A&M University, College Station, TX, (19)Dept. of Ecosystem Science and Management, Texas A&M University, College Station, TX

Conifers are evolutionarily ancient group of trees that form the dominant vegetation throughout the high latitude boreal forests as well as some lower latitude temperate forests of North America, Europe, and Asia. They possess exceptional longevity and play an integral economic and ecological role across much of the world. Here we investigate de novo assembled transcriptomes of four white pine species: western white pine (*Pinus monticola*), limber pine (*Pinus flexilis*), whitebark pine (*Pinus albicaulis)*, and sugar pine (*Pinus lambertiana*). All are five-needle white pines classified as members of the subgenus Strobus. Among them *Pinus flexilis*, and *Pinus albicaulis* are high elevation species and will be largely affected by climate change. Annotation of the four white pine needle leaf tissue transcriptomes revealed a core set of 2025 gene families shared across the four white pines among which 408 showed signatures of selection. Of these 408, 39 are estimated to be under positive selection (average dN/dS > 1), 9 are under neutral selection (average dN/dS < 1 and average dN/dS > 0.95), and the remaining 360 are under purifying selection (average dN/dS <0.95). Among the positively selected genes FBK-Skip6, E3 Ubiquitin Protein Ligase and UTP7 will be highlighted. A comparative study examining the gene space among the four species including candidate genes involved in drought tolerance, disease and pest resistance and phenology is underway. The PineRefSeq project which has produced genome assemblies for loblolly pine, sugar pine, and Douglas-fir has been annotated with the assistance of deep transcriptomes of this study. Adaptive evolution and patterns of selection among the four white pines will advance our understanding of this ancient non-model plant group.

W741: Population and Conservation Genomics

Whole Genome Sequencing of California Condors is Now Utilized for Guiding Genetic Management

Oliver Ryder[1], Webb Miller[2], Katherine Ralls[3], Jonathan D. Ballou[4], **Cynthia C. Steiner**[1], Anna Mitelberg[1], Michael Romanov[5], Leona G. Chemnick[1], Michael Mace[6] and Stephan Schuster[7], (1)San Diego Zoo Institute for Conservation Research, Escondido, CA, (2)Pennsylvania State University, Pennsylvania, PA, (3)Smithsonian Conservation Biology Institute, Washington, DC, (4)Center for Conservation and Evolutionary Genetics, Washington, DC, (5)University of Kent, Canterbury CT2, United Kingdom, (6)San Diego Zoo Safari Park, San Diego, CA, (7)Pennsylvania State University, University Park, PA

The California condor is a critically endangered avian species that, in 1982, became extinct in the wild. Its survival has persevered through a captive breeding program and reintroduction efforts within its former range. As of April, 2015, 421 California condors, including 204 flying in the wild constituted the extant population. Concern regarding preservation of genetic diversity and inbreeding, have led to intensive population management supported by molecular genetics research and, more recently, the application of genomic methodologies. 36 complete California condor genomes, representing the whole gene pool of the species, have been sequenced identifying about 4 millions polymorphic sites (SNPs). This has allowed reassessment of kinship among the founder birds, which is now being applied to selecting breeding pairs for the ongoing

captive propagation effort. A genetic disease, chondrodystrophy, is inherited consistent with an autosomal recessive mode of transmission in condors. Utilizing whole genome sequencing of affected chicks and their carrier parents, a series of linked markers localized in a 1 Mb region of the condor genome have been employed to detect carrier condors heterozygous for the lethal mutation. This information can be incorporated into population management to reduce the risk of reproductive failure, as reintroduced populations begin to expand.

W742: Population and Conservation Genomics

Rapid Evolutionary Response to Infectious Cancer

Sarah Hendricks[1], Brendan Epstein[2], Barbara Schonfeld[3], Menna Jones[3], Andrew Storfer[2] and Paul Hohenlohe[1], (1)University of Idaho, Moscow, ID, (2)Washington State University, Pullman, WA, (3)University of Tasmania, Hobart, Australia

Tasmanian Devil Facial Tumor Disease (DFTD) is a unique infectious cancer that threatens the persistence of Tasmanian devils (*Sarcophilus harrisii*). Tumor cells are transmitted by biting among reproductively mature individuals, and infected individuals rarely reproduce more than once before succumbing to the disease. However, long-infected populations have persisted, and there is recent evidence for population recovery and tumor regression in a few individuals. Understanding the genetic basis of resistance to DFTD could illuminate mechanisms of cancer resistance and potentially allow prediction of the future course of DFTD. We have used Restriction-site Associated DNA sequencing (RADseq) to scan the genome for candidate loci exhibiting a response to selection by DFTD, and to determine geographic structure of genetic variation, both across the genome and at candidate loci. We identified genomic regions with concordant signatures of selection in response to DFTD across three focal populations. These results suggest rapid evolution in response to strong selection imposed by DFTD, and the potential for genetic variation for resistance to the disease that could facilitate population recovery, despite the overall low levels of genetic diversity within the species. Our phylogeographic analysis confirms previous evidence for two major clusters of devil populations roughly divided into the eastern two-thirds and western third of Tasmania. The only remaining uninfected populations are in the northwest corner. Our results suggest that genetic distinctiveness in western devil populations may have implications for the future spread of the disease.

W743: Population and Conservation Genomics

Genomic Variation and Population Structure of the Threatened Neosho Madtom (*Noturus placidus*)

Lynsey K. Whitacre[1], Mark L. Wildhaber[2], Gary S. Johnson[3], Justin M. Downs[4], Tendai Mhlanga-Mutangadura[3], Vernon M. Tabor[5], Daniel Fenner[6] and **Jared E. Decker**[7], (1)Informatics Institute, University of Missouri, Columbia, MO, (2)U.S. Geological Survey; Columbia Environmental Research Center, Columbia, MO, (3)Department of Vet Pathobiology, University of Missouri, Columbia, MO, (4)Peoria Tribe, Miami, OK, (5)U.S. Fish and Wildlife Service Kansas Ecological Services Field Office, Manhattan, KS, (6)U.S. Fish and Wildlife Service Oklahoma Ecological Services, Tulsa, OK, (7)Division of Animal Sciences, University of Missouri, Columbia, MO

The Neosho madtom (*Noturus placidus*) is a small catfish, generally less than 3 inches in length, unique to the Neosho-Spring River system within the Arkansas River Basin. It was federally listed as threatened in 1990, largely due to habitat loss. As part of conservation efforts, we generated whole genome Illumina paired-end sequence data from ten Neosho madtom (average 39X coverage) originating from three geographically separated subpopulations to evaluate genetic diversity and population structure. One slender madtom (*Noturus exilis*) was also sequenced as an outgroup. Lack of a reference genome necessitated variants be discovered using De Bruijn graphs implemented in CORTEX v1.0.5.21. Approximately 1.64 million high confidence single nucleotide polymorphisms (SNPs) were observed. Only 86,155 SNPs were variable across Neosho madtoms sequenced, indicating overall low level genetic diversity. While principal component analysis based on these genotypes accurately clustered individuals from the same location together, insignificant eigenvalues indicated weak population structure, suggesting these subpopulations are genetically compatible for reintroduction among these three locations. We also completed a draft *de novo* assembly of the Neosho madtom genome from 120X of sequences pooled across 3 individuals. We assembled the ~1 Gb genome into 149,885 contigs with a N50 of 12,261 bp, demonstrating value in assembling a genome from a population that is closely related to a species of economic interest (i.e., channel catfish, *Ictalurus punctatus*) but has lower genetic diversity and is easier to assemble. Ongoing efforts aim to improve the assembly by sequencing mate pair libraries and to estimate historical effective population sizes.

W744: Population and Conservation Genomics

A Comprehensive Exome Scan for Signatures of Rapid Fisheries-induced Evolution

Nina Overgaard Therkildsen, Cornell University, Ithaca, NY, Stephan B. Munch, NOAA Southwest Fisheries Science Center, Santa Cruz, CA, David O. Conover, Stony Brook University, Stony Brook, NY and Stephen R. Palumbi, Stanford University, Pacific Grove, CA

It has become evident that fisheries can inflict evolutionary changes in the exploited populations. Yet, so far almost nothing is known about the underlying genomic basis for the widespread changes in life history traits observed across the world's fish stocks: What types of genetic variation does fisheries-selection act on, how extensively does it impact the genome, and how reversible are the changes once fishing stops? To address these questions, we have returned to a seminal experiment that demonstrated substantial evolution in growth rates in response to size-selective fishing over just five generations in the Atlantic silverside. We use low coverage 'in silico' exome capture for 876 individually barcoded fish to track allele frequency changes in different experimental populations at >800,000 SNPs covering ~80% of the exome. We have identified strong differentiation between large- and small-selected populations at 1,118 SNPs and nearly fixed differences in 17 genes, including bone morphogenic proteins and macrophage stimulating factors, previously linked to growth. However, selected genes maintain high overall levels of diversity, indicating that selection has primarily acted as 'soft' sweeps on old alleles. These findings improve our mechanistic understanding of fisheries-induced evolution and provide an important baseline for assessing genomic changes in wild fish populations.

W745: Population and Conservation Genomics

Genomic Flatlining in the Endangered Island Fox (*Urocyon littoralis*)

Jacqueline Robinson[1], Diego Ortega Del Vecchyo[1], Zhenxin Fan[2], Bridgett Vonholdt[3], Bernard Kim[1], Clare Marsden[1], Kirk E. Lohmueller[1] and Robert K. Wayne[1], (1)University of California, Los Angeles, Los Angeles, CA, (2)Sichuan University, Chengdu, China, (3)Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ
With species declining worldwide due to exploitation and habitat loss, a vital element in mitigating future biodiversity loss will include understanding the genetic consequences of small population size and isolation. Population genetics theory holds that small populations in isolation will suffer from declining diversity, increasing genetic load, and eventual extirpation, in a process dubbed the “extinction vortex.” Island species therefore present an ideal system for empirically studying the impacts of small population size and isolation. Island foxes (*Urocyon littoralis*) are dwarfed descendants of the mainland gray fox (*U. cinereoargenteus*) that have persisted on California’s Channel Islands for thousands of generations. We sequenced complete genomes from each island population and the mainland to examine heterozygosity and deleterious variation within individual genomes. The genome of the San Nicolas Island fox (*U. l. dickeyi*) is almost entirely monomorphic, and retains the lowest observed genomic heterozygosity of any outbred species to date. Furthermore, genomes from all island populations exhibit reduced variation and an elevation in the proportion of deleterious alleles, resulting from strong genetic drift and reduced efficacy of selection. Through demographic simulations, we show that the observed patterns of heterozygosity within the island genomes can be attributed to severe bottlenecks and long-term small population size. The case of the island fox, with its persistence despite extreme lack of genetic variability and increased genetic load, raises questions about the generality of the small population paradigm.

W746: Population and Conservation Genomics

Understanding the Origin of Genetic Variations in Rare Breeds: the Importance of Gene-Flow and Demography

Laurent Frantz, Wageningen University, Wageningen, Netherlands
Domestic populations display a myriad of phenotypic variations. In farm animals, many of these variations are found in local/rare breeds. However, these populations are under threat due to the predominance of few breeds in industrial settings. Understanding the mechanisms that govern such phenotypic diversity is crucial to inform conservation efforts. To investigate this issue we sequenced over 100 pig genomes from diverse rare or commercial breeds as well as multiple populations of wild boars. We used this data to model the domestication history of pigs. Our results strongly support models involving gene flow between wild and domestic forms throughout history. Gene flow was high from wild to domestics but low the other way around. We also show that wild boars have undergone massive bottlenecks since domestication resulting in low effective population size. On the other hand, gene-flow have increased the effective population size of domestic pigs. Moreover, we show that besides for a few portions of the genome that are homogeneous among breeds due the their importance for maintaining domestic traits, genetic variations segregating between pig breeds were likely acquired through gene-flow from diverse local wild populations. Our complex picture of the domestication history of pigs demonstrates that domestic or wild boars should not be considered as genetically homogeneous groups. Instead, we argue that rare/local breeds can be seen as reservoirs for genetic variations that have been lost in the wild due to bottlenecks and should therefore be the focus of conservation efforts.

W747: Post-transcriptional Gene Regulation

Large Scale Identification and Characterization of RNA Binding Protein-RNA Networks

Gene W. Yeo, University of California, San Diego, La Jolla, CA
RNA binding proteins (RBPs) play essential roles in cellular physiology by interacting with target RNA substrates. Elucidating the binding sites of RBPs is a crucial first step in deepening our understanding of RNA biology. I will present our ongoing efforts as part of the ENCODE consortium to define the binding sites at nucleotide-level resolution for hundreds of RNA binding proteins.

W748: Post-transcriptional Gene Regulation

Differential Intron Retention is a Key Event of Alternative Splicing-Driven Transcriptome Adaptation to Environmental Stresses in *Populus trichocarpa*

Sergei Filichkin[1], Michael Hamilton[2], Palitha Dharmawardhana[3], Anireddy S. Reddy[2], Asa Ben-Hur[2] and Pankaj Jaiswal[3], (1)Oregon State University, Corvallis, OR, (2)Colorado State University, Fort Collins, CO, (3)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR
Alternative splicing (AS) of precursor messenger RNAs (pre-mRNAs) is widespread across metazoan organisms. In plants the majority of intron-containing genes (42-61%) are alternatively spliced with intron retention (IR) being the prevalent class. The repertoire of AS can be modulated by various environmental stresses. We investigated the extent of abiotic stress-driven differential IR (DIR) events across principal tissue types of western poplar (*Populus trichcarpa*). Osmotic (dehydration or high salinity) or temperature (heat or cold) stresses triggered a broad range of DIR events in leaf, root, and xylem tissues. Each stress treatment induced a distinct set of stress-specific DIR events. The majority of transcripts harboring DIR events were unique between short-term and prolonged treatment phases. However, several subsets of common DIR events were ubiquitously induced by two or more stress treatments. Among transcripts harboring multiple stress-regulated retained introns some DIR events occurred with high frequency across all treatments and tissues. Such highly responsive events were often observed in transcripts derived from genes regulating stress response. Stress-induced DIR events were present across key gene families regulating splicing, stress-response, plant development, cell wall metabolism, transcription, and circadian rhythms.

W749: Post-transcriptional Gene Regulation

Genome-Wide Landscape of Alternative Polyadenylation in Rice: Impact on Developmental Gene Expression Regulation and QTL

Qingshun Quinn Li[1,2], Haihui Fu[2], Dewei Yang[3], Xiaohui Wu[2], Guoli Ji[2] and Xinfu Ye[3], (1)Western University of Health Sciences, Pomona, CA, (2)Xiamen University, Xiamen, China, (3)Fujian Agricultural Science Academy, Fuzhou, China
Alternative polyadenylation (APA), a phenomenon that a transcript uses one of multiple poly(A) sites to produce a defined 3’-end and a poly(A) tail, has been recognized as a wide-spread regulatory mechanism in eukaryotic gene expression. Using a Poly(A)-Tag sequencing (PAT-seq)

protocol, we compared APA in 14 different tissues/developmental stages of rice. These include different tissues roots and leaves, and different stages in embryo, endosperm, imbibed seeds and flower parts (i.e. husk, mature pollen, anther, and pistil). A total of 68,220 poly(A) sites on 13,419 APA genes (~48% of total expressed genes) were identified. We also found that many important function genes have APA switching events between different developmental stages or different tissues, e.g., a rubisco small subunit gene associated with photosynthesis has an APA switching pattern in leaves between seedling and tillering stages. Poly(A) sites across pollen are unique when compared to all other tissues, with the most tissue-specific poly(A) sites and very different poly(A) signals of the intronic poly(A) sites. Using statistical analysis, it was found that Quantitative Trait Loci (QTL) tend to have APA genes. In addition, we found that many APA genes of high expression levels were mainly distributed in three category QTL, yield, vigor, and anatomy, further suggesting that APA gene may play an essential role in the determination of agronomic traits in rice. Besides aiding rice genome annotation, these data provide the most comprehensive and highest resolution APA information in plants for further interrogation of the role of APA in plants.

W750: Post-transcriptional Gene Regulation

Global Analyses of Light-Regulated Alternative Splicing in Plants

Chueh-Ju Shih, Hshin-Ping Wu, Wen-Dar Lin and **Shih-Long Tu**, Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Light is one of the most important factors influencing plant growth and development. Changes of light condition dramatically affect developmental programs throughout the life cycle of plants. Light-sensing photoreceptors play major roles in regulating these processes via signal transduction and gene regulation. Although various levels of gene expression are known to be modulated by photoreceptors, regulation at the pre-mRNA splicing step is less discussed. We have performed mRNA sequencing to analyze transcriptome changes during light exposure in the moss *Physcomitrella patens*. Our data showed that light induces intensive alternative splicing. Light-responsive intron retention preferentially occurred in transcripts for photosynthesis and translation, which reveals that light-mediated splicing regulation has transcript specificity. Moreover, intron retention was rapidly induced by light but misregulated in photoreceptor-deficient and -knockout mutants, suggesting the involvement of photoreceptors in splicing regulation. Preliminary data indicate the red/far-red photoreceptor phytochrome physically interacts with splicing regulators in the nucleus in a red-light dependent manner. We therefore propose that when plants expose to light, alternative splicing is rapidly fine-tuned to re-organize certain metabolic processes and regulate gene expression. Photoreceptors directly participate in regulation of alternative splicing through interacting with regulatory proteins to modulate splicing activity.

W751: Post-transcriptional Gene Regulation

ORF Discovery Using High-Resolution Ribosome Profiling Data in Arabidopsis

Polly Yingshan Hsu[1], Lorenzo Calviello[2], Larry H. Wu[3], Evan W. McConnell[4], Leslie M. Hicks[4], Uwe Ohler[2] and Philip N. Benfey[5], (1)Duke University, Durham, NC, (2)Max Delbrück Center, Berlin, Germany, (3)North Carolina State University, Raleigh, NC, (4)UNC, Chapel Hill, Chapel Hill, NC, (5)Duke University and Howard Hughes Medical Institute, Durham, NC

Identifying protein-coding genes is essential as proteins are usually the ultimate players in the cell. Recently, ribosome profiling (deep sequencing of ribosome footprints) has revealed that ribosomes occupy many regions in the transcriptome outside of the annotated protein-coding sequences, such as 5'UTRs and non-coding RNAs. To distinguish real translation events from contaminants in ribosome purification, we developed an efficient ribosome profiling protocol in Arabidopsis, which allows one to precisely infer codons being translated within ribosome footprints. Subsequently the footprint data are subjected to a modified Fourier Transform method to statistically test whether the periodicity of footprints of a given ORF reflects translocation patterns of actively translating ribosomes. With stringent criteria, ORFs from over 18,000 genes have been identified that have excellent agreement with the current annotation. In addition, several hundred upstream open reading frames (uORF) within 5'UTRs as well as a few dozen novel ORFs from transcripts annotated as non-coding have been identified with high coverage and stringent statistics. Our approach not only provides experimental support for predicted gene models but also discovery of previously unannotated ORFs. As the Arabidopsis genome is among the best annotated, we expect this approach will be even more powerful when applied to other species.

W752: Poultry 1

Gene Expression Analysis Using 3' mRNA Sequencing Reveals Exciting Prospect for Functional Genomic Research at Population-Level: Application in a Chicken Study

Behnam Abasht, Weixuan Fu and Zhu Zhuo, Department of Animal Science, University of Delaware, Newark, DE

RNA sequencing (RNA-seq) has revolutionized the study of gene expression in animals, plants and microorganisms. However, because of its high cost, this technology has been mainly used in experiments with limited number of samples. To examine a cost-effective alternative, we used a method, which confines sequencing to the 3'-end of mRNA and produces just one fragment per transcript, resulting in a dramatic decrease in sequencing cost. Total RNA isolated from chicken adipose tissue samples was used for cDNA library preparation using QuantSeq 3'mRNA-seq library Prep Kit. Sixty-one uniquely indexed cDNA libraries were pooled and sequenced on one lane on the Illumina Hiseq 2500. On average, 2.24 million reads per sample were generated, 90.1% of which were mapped to the chicken reference genome (Ensembl Galgal4). For more than 70% of the genes with detectable expression, we redefined the 3'-end and identified alternative polyadenylation sites within the 3'-untranslated regions. To compare gene expression measures between 3'-RNA-seq and RNA-seq technologies, we used data from a subset of 20 samples that were previously used in a RNA-seq study of feed efficiency. The correlation of the log10(fold-change) for gene expression (high- *vs.* low-feed efficiency birds) between these two methods was 0.90. In conclusion, 3'-RNA-seq is a cost effective method amenable to global gene expression studies at population-level, e.g., expression QTL (eQTL) mapping. Also, it allows for accurate detection of the 3'-end of transcripts, enabling verification of the current gene model annotations and global characterization of alternative polyadenylation.

W753: Poultry 1

Molecular Mechanisms Associated with Dietary Methionine Deficiency in Meat-Type Chickens

Samuel E. Aggrey, Fernando Gonzalez-Ceron and Romdhane Rekaya, University of Georgia, Athens, GA

Methionine is the first limiting amino acid in a typical poultry diet. Restriction of dietary sulfur amino acids (SAA) to growing chickens leads to reduced growth and oxidative stress. The molecular mechanisms that underlie such restrictions remain scant. We studied the molecular mechanisms that underlie SAA restrictions in broiler chickens from 3-5 weeks of age in the *Pectoralis major* major muscle using next generation sequencing. Fold change of ≥1.5 and false discovery rate of ≤0.05 were used as criteria for differentially expressed genes (DEGs). There were 554 downregulated DEGs and 365 unregulated DEGs in the deficient group compared to the controls. Signaling pathway impact analysis showed that the Fc gamma R-mediated phagocytosis, NK cell mediated cytotoxicity and NF-kappa B signaling pathways were activated whereas the B cell receptor signaling pathway was inhibited in the *Pectoralis major* of the SAA restricted birds. The current study suggests that restriction in dietary methionine in growing chickens leads to cytoskeleton modification that reduces cell to cell communication, potential increase in oxidative stress and abnormal methylation which induces inflammation response. Birds eliciting such inflammation response may compromise their immune system and their ability to fight infections.

W754: Poultry 1

Potential Driver Mutations for Marek's Disease

Hans Cheng, USDA, ARS, ADOL, East Lansing, MI and Alec Steep, Michigan State University, East Lansing, MI

Marek's disease (MD), a lymphoproliferative disease of chickens caused by the highly pathogenic Marek's disease virus (MDV) is the most serious chronic disease problem that costs the worldwide poultry industry $1-2 billion per year. In addition to its agronomic relevance, MD is a valuable biomedical model for viral-induced transformation. MDV Meq is a bZIP transcription factor that is necessary but not sufficient for the induction of lymphomas. To address whether genomic alterations are necessary for MDV-induced transformation, 26 MD tumors have been both DNA and RNA sequenced to identify somatic mutations. Currently progress of this effort will be discussed. The resulting information combined with ongoing experiments should result in a significant increase in knowledge on how MD vaccines protect birds as well as genes and genetic markers for genomic selection for enhanced MD genetic resistance in commercial flocks.

W755: Poultry 1

HPIDB: A Curated Database for Host-Pathogen Interactions

Mais G. Ammari[1], Cathy R. Gresham[2], Fiona M. McCarthy[3] and **Bindu Nanduri**[2,4], (1)School of Animal and Comparative Biomedical Sciences, Tucson, AZ, (2)Institute for Genomics, Biocomputing & Biotechnology, Mississippi State, MS, (3)University of Arizona, Tucson, AZ, (4)College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Understanding the interplay between host and pathogen that underpins health/disease enables identification of potential targets for therapeutic, prophylactic and intervention strategies to eliminate or reduce the severity and economic impact of infectious diseases. However, existing host-pathogen interaction (HPI) data is generally poorly represented in molecular interaction databases or is predicted based on sequence analyses. Species-specific HPI data for agricultural animals including chicken is very limited. To provide agricultural researchers with HPI information, we developed Host-Pathogen Interaction Database (HPIDB). HPIDB (http://www.agbase.msstate.edu/hpi/main.html) contains 43,276 manually curated entries in the current release. Since the first description of the database in 2010, we made multiple enhancements to HPIDB data and interface services. Notably, HPIDB now provides targeted biocuration of molecular interaction data. As a member of the International Molecular Exchange consortium, annotations provided by HPIDB curators meet community standards to provide detailed contextual experimental information and facilitate data sharing. Moreover, HPIDB provides access to rapid community annotations that capture minimum information regarding molecular interactions to address immediate researcher needs for HPI network analysis. In addition to curation, HPIDB integrate HPI from existing external sources and contain tools to infer additional HPI where annotated data is scarce. Compared to other molecular interaction databases, our data collection approach ensures HPIDB users access the most comprehensive HPI data from a wide range of pathogens and their hosts (currently, 567 pathogen and 68 host species). Improvements to user interface include enhanced search capacity, addition of Gene Ontology functional information, and implementation of network visualization.

W756: Poultry 1

***Gallus gallus* Data at NCBI: From Genome Annotation to Gene Curation**

Kim D. Pruitt[1], David Webb[1], Terence D. Murphy[2] and Francoise Thibaud-Nissen[1], (1)National Center for Biotechnology Information (NCBI/NLM/NIH), Bethesda, MD, (2)National Center for Biotechnology Information, NLM, NIH, Bethesda, MD

The National Center for Biotechnology Information (NCBI) reference sequence (RefSeq) project provides several resources to support the chicken research community. These include: genome annotation; evidence-based transcript and protein sequence records; BLAST databases; annotation reports; FTP data; and entries in NCBI's Gene resource. Gene, transcript, and protein data are provided using a combined approach of targeted manual curation, collaboration with the AgBase and the Chicken Gene Nomenclature Committee (CGNC), and whole genome annotation by NCBI's eukaryotic genome annotation pipeline. The eukaryotic genome annotation pipeline uses the available known RefSeqs (NM_, NR_, and NP_ accessions that have curation support), RNA-Seq, cDNA/EST/TSA, and protein alignments to generate whole genome annotation and corresponding model RefSeq RNA and protein sequence products (XM_, XR_, XP_ accessions). Manual curation focuses on both sequence and gene data. For example, in 2015 curators reviewed putative sequence errors (such as frameshifts), removing redundant gene and ortholog data, and nomenclature conflicts affecting paralogs, gene families, or between NCBI's Gene database and CGNC. The presentation will present genome annotation results for the current public version of the chicken assembly, examples of RefSeq curation, and information on data access. More information about NCBI's RefSeq project is available from: www.ncbi.nlm.nih.gov/refseq/

W757: Poultry 1

Hepatic Gene Expression and Ovarian Follicle Development in Broiler Breeder Hens

Patricia A. Johnson, Cornell University, Ithaca, NY

Hepatic gene expression and ovarian follicle development in broiler breeder hens.

W758: Poultry 1

Update on the Most Recent Chicken Genome Assembly

W759: Poultry 1

Long-Non Coding RNAs Repertoire in Liver and Adipose Tissue in Chicken

Kevin Muret[1], Christophe Klopp[2], Valentin Wucher[3], Diane Esquerre[4], Fabrice Legeai[5], Frederic Lecerf[6], Colette Desert[6], Morgane Boutin[6], Herve Acloque[7], Elisabetta Giuffra[8], Sarah Djebali[9], Sylvain Foissac[10], Thomas Derrien[3] and **Sandrine Lagarrigue**[6], (1)Agrocampus Ouest - INRA, UMR1348 PEGASE, RENNES, France, (2)INRA - SIGENAE, Castanet Tolosan, France, (3)University of Rennes I, RENNES, France, (4)Plateforme Genomique, Castanet tolosan, France, (5)INRA - UMR IGEEP, RENNES, France, (6)Agrocampus Ouest - INRA, UMR1348 PEGASE, Rennes, France, (7)INRA-GenPhySE, Castanet-Tolosan, France, (8)INRA, UMR de Génétique Animale et Biologie Intégrative, Jouy-en-Josas, France, (9)INRA, GenPhySE, Castanet Tolosan, France, (10)INRA-GenPhySE, Castanet Tolosan, France

Improving the functional annotation of the chicken genome is a key challenge in bridging the gaps between genotype and phenotype. Among all transcribed regions, long non-coding RNAs (lncRNAs) are a major component of the transcriptome and the use of whole transcriptome sequencing (RNA-seq) has greatly improved the identification and characterization of these non-coding genes.

In this study, we focused on liver and adipose tissues because of their importance in various economically traits in which energy storage and mobilisation play a roles and also due to the high cell homogeneity of these tissues. We thus investigated 16 RNAseq experiments from each tissue with strand-specific reads for identifying lncRNAs. We automated and implemented a program, called FEELnc, which allowed us to identify around 3,000 chicken lncRNAs longer than 200 bp, multi and mono-exonic and without protein-coding capabilities. FEELnc also classified lncRNAs based on their genomic localizations with respect to the ENSEMBL protein-coding annotation.

The intergenic lncRNAs class (~90%) was characterized with respect to the distance and orientation with the closest mRNAs and the intragenic lncRNAs class (~10%) was extracted based on their overlap with mRNAs exons and introns. We then characterized more deeply this repertoire in terms of structure and expression and compare these features with the latest human lncRNA repertoire (Derrien et al 2012). In particular, we identified tissue-specific lncRNAs, and one lncRNA as a good candidate to the regulation of lipid metabolism in liver. This study will further be extended to more species and tissues/cell lines in the context of the FAANG project “Fr-AgENCODE”.

W760: Poultry 1

Proteogenomic Integration for Mitochondrial Dysfunction in Chicken DF-1 and LMH Immortal Cell Lines

Byung-Whi Kong, Dongwon Seo, Bhuwan Khatri, Ayla Smartt, Sami Dridi and Walter Bottje, University of Arkansas, Fayetteville, AR

Continuously growing immortal chicken cell lines including immortal DF-1 chicken embryo fibroblast (CEF) and LMH (leghorn male hepatoma) have been widely used for research areas of virology, cell biology, and molecular biology in avian species. The DF-1 line, which was derived spontaneously from primary CEF cells, and the LMH line, which was chemically induced, have been widely used for the propagation of avian infectious viruses including avian influenza, avian infectious bronchitis virus, Marek’s disease virus serotype 1 (MDV-1), avian metapneumovirus (AMPV), and infectious laryngotracheitis virus (ILTV). Earlier studies with primary cells, DF-1 and LMH cell lines showed various differences in the mechanisms of cell cycle regulation, telomerase activities and mitochondrial functions in response to oxidative stresses. Using RNAseq, shotgun proteomics and bioinformatic pathway analysis, functionally integrated pathways were characterized in primary CEF, DF-1 CEF and LMH cells to regulate various cellular mechanisms including mitochondrial dysfunctions in this study.

W761: Poultry 1

The Role of the Avian Vasotocin 1a (V1a) and V1b Receptors in the Neuroendocrine Regulation of Stress

Wayne Kuenzel[1], Alexander Jurkevich[2], Gurueswar Nagarajan[1], Nguiessan Alphonse Aman[3] and Seong W. Kang[1], (1)University of Arkansas, Fayetteville, AR, (2)University of Missouri-Columbia, Columbia, MO, (3)School of Medicine, Alassane Quattara University, Bouake, Ivory Coast

The avian vasotocin receptors type 4 (VT4R) and 2 (VT2R) are homologous to the mammalian vasopressin receptors V1aR and V1bR, respectively. With respect to the neuroendocrine regulation of stress in birds, it has been shown following acute psychological stress, that the avian V1aR and V1bR and corticotropin releasing hormone 1 receptor (CRH-R1) and 2 receptor (CRH-R2) displayed significantly different gene expression levels compared to controls. Brain distribution of the avian V1aR showed its presence not only in neurons but also in glia, particularly in all ten circumventricular organs. Dense V1aR immunoreactivity (ir) was found in the organum vasculosum of the lamina terminalis (OVLT) and sub-fornical organ (SFO). In birds subjected to a physical stressor, intraperitoneal injection of sodium chloride, significantly increased V1aR gene expression occurred in the SFO as well as the OVLT compared to controls. It was also shown that the dorsal portion of the nucleus of the hippocampal commissure (nHpC) contains a population of CRH neurons in close vicinity to V1aR ir glial cells. Cells within the dorsal nHpC likewise showed significant activation following acute stress based upon Fos protein immunoreactivity. Functional data suggest that (1) both the avian V1aR and V1bR are involved in the neuroendocrine regulation of psychological stress, (2) the avian SFO and OVLT containing V1aR-ir glia appear to play a role in physical stress and (3) CRH neurons within the dorsal region of the nHpC are involved in the stress response. Supported in part by grants from NSF (#O842937) and the Arkansas Biosciences Institute.

W762: Poultry 1

Effects of Thermal Challenge on Turkey Muscle Development and Meat Quality

Gale M. Strasburg[1], Daniel L. Clark[2], Cindy Coy[2], Sandra G. Velleman[2], Galen George[1] and Kent M. Reed[3], (1)Michigan State University, East Lansing, MI, (2)The Ohio State University, Wooster, OH, (3)University of Minnesota, St. Paul, MN

Exposure of newly hatched turkey poults to hot or cold thermal stress often results in detrimental effects on breast muscle growth and development. Typical changes include increased lipid deposition and damage to muscle ultrastructure, leading to inferior meat quality with

consequent economic losses to producers and processors. A leading candidate for mediating the effects of thermal challenge on muscle growth and development is the satellite cell which is the progenitor of post-hatch muscle growth. We hypothesized that hot or cold thermal challenge alters gene expression, resulting in changes in muscle structure that ultimately produce inferior meat quality characteristics in the market-age bird. This study examined effects of temperature on proliferation and differentiation of cultured turkey p. major satellite cells from the RBC2 (control) line and growth-selected F line. While cells from both lines show temperature-dependent changes in function, the F line cells are more sensitive than RBC2 satellite cells. In a subsequent study, newly hatched RBC2 and F line turkey poults were exposed to different brooding temperatures (31, 35, or 39C) for 3 days followed growth to 16 weeks under standard conditions. Analysis of p. major ultrastructure indicates larger muscle fiber width and reduced number of cells per unit area in F line birds at the elevated brooding temperatures. Understanding the mechanism of satellite cell response to thermal challenge will be useful in development of adaptation strategies to improve thermo-tolerance of the birds as well as meat quality characteristics.

W763: Poultry 1
Functional Annotation of

W764: Poultry 1
Genomics/Metabolomics of

W765: Poultry 1
Genetics of Dilated Cardiomyopathy in the Turkey
Edward Smith and Kwaku Gyenai, Virginia Tech, Blacksburg, VA
Kwaku Gyenai, Jun Xu, and Ed Smith, Comparative Genomics Lab, department of Animal and Poultry Sciences, Virginia Tech, Blacksburg, VA 24060
Dilated cardiomyopathy is reported to account for about 2% of mortality in turkeys. . In several studies, we investigated the effects of nutritional and genetic factors on the incidence and severity of toxin-induced dilated cardiomyopathy in the domestic turkey, *Melagris gallopavo*. In one study, the results of which will be presented here, we evaluated the effect of dietary vitamin E and selenium (Se) on oxidative stress (OS) and the incidence and severity of furazolidone (Fz)-induced dilated cardiomyopathy. Standard turkey diets were supplemented with different concentrations and combinations of vitamin E (0, 50 and 100 IU/kg) and selenium (0.0, 0.3 and 0.5 mg/kg). Malondialdehyde (MDA), glutathione (GSH), glutathione peroxidase (GPx), and plasma uric acid (PUA) were used as biomarkers for oxidative stress. At two wks of age, measurements of MDA for birds fed normal diet were not different from those fed diet with toxic levels of Fz. However, GPx was increased for birds fed diets supplemented with 0.5 mg/kg Se, combinations of 50 IU/kg vitamin E and 0.5 mg/kg Se, and 100 IU/kg vitamin E and 0.5 mg/kg Se. GPx was highest for birds fed Fz-containing diets and a combination of 100 IU/kg vitamin and 0.5 mg/kg Se. No differences were observed for poults (young turkeys) fed normal diets PUA; however, PUA was high for poults fed Fz-containing. Among birds fed Fz-containing diets, reduced MDA levels were observed for birds fed diets with a combination 50 IU/kg vitamin E and 0.3 mg/kg Se, and 0.5 mg/kg Se, respectively. At two weeks of age, poults fed normal diets supplemented with 100 IU/kg and 0.5 mg/kg vitamin E and Se combined had the highest GSH. Among birds fed Fz-containing diets, GSH was highest for those fed a 100 IU/kg vitamin E. At four weeks of age, birds fed normal diets and combinations of 50 IU/kg vitamin E and 0.5 mg/kg Se, 100 IU/kg vitamin E had the lowest concentration of MDA but highest GPx and GSH. Though results of the biomarkers were inconsistent from two to four wks of age, for turkeys fed diet containing different concentrations and combinations of dietary antioxidant, our current findings suggest a role for increased OS in the progression of DCM.

W766: Poultry 1
Update on NRSP8 Bioinformatics

W767: Poultry 1
Transcriptional Analysis of Rathke's Pouch Formation during Chick Embryonic Development by RNAseq
Tom E. Porter[1], Monika Proszkowiec-Weglarz[1] and Hsiao-Ching Liu[2], (1)University of Maryland, College Park, MD, (2)North Carolina State University, Raleigh, NC
The anterior pituitary gland plays an essential role in the regulation of many physiological processes such as growth, reproduction, lactation, stress, and metabolism in vertebrates. Formation of Rathke’s pouch, the precursor of the anterior pituitary gland, is a multi-step process regulated by cell interactions, signaling pathways, and transcription factors, that starts with evagination of the oral ectoderm. Previously, we used laser capture microdissection, RNA amplification, and reverse transcription quantitative real-time PCR (RT-qPCR) to characterize mRNA levels for selected genes in the oral ectoderm and adjacent neural ectoderm during Rathke's pouch formation in chicken embryos (embryonic day 2.5 to 7). However, each of the selected genes analyzed had been implicated previously in pituitary development. In the current study, RNAseq was performed on these samples to analyze changes in mRNA levels on a genome-wide scale and identify novel genes associated with Rathke's pouch formation and pituitary development.

W768: Poultry 1
***In vitro* Insights into the Transcriptomic Response of Chicken Cells to Stressors**
Susan J. Lamont[1], Angelica Van Goor[1], Anna Slawinska[1,2], John Hsieh[1] and Carl J. Schmidt[3], (1)Iowa State University, Ames, IA, (2)University of Technology and Life Sciences, Bydgoszcz, Poland, (3)University of Delaware, Newark, DE
In genomic studies, scales of investigation can range from small (molecular and cellular) to large (organ, organism, population and system-biology), with each type uniquely contributing toward a more comprehensive understanding of response plasticity in the face of environmental perturbations. We used isolated cells to characterize transcriptomic responses to the single or multiple stressors of high environmental temperature and lipopolysaccharide (LPS)-induced endotoxemia, in vitro. Endotoxemia in animals can result from heat stress-induced “leaky gut”. Cell types included the chicken HD11 macrophage-like cell line, and bone marrow derived antigen presenting cells (BM-APC) of highly

inbred Fayoumi and Leghorn chicken lines. Gene expression analysis was conducted using a panel of heat shock protein (HSP), stress-related, signaling molecule and immune-response genes. The HD11 cell line responded to high temperature with increased mRNA levels of several HSP and co-chaperone genes. Up-regulation of effector cytokine genes in response to LPS was greater in HD11 cells in the heat-stress environment than in thermoneutral conditions. In BM-APC of Fayoumi and Leghorn, several HSPs were strongly induced with heat treatment, with few differences between lines. Many immune-related genes in BM-APC responded strongly to one or more treatments with either up- or down-regulation, and most were significantly different between Leghorn and Fayoumi. Fayoumi BM-APC produced more nitric oxide and had higher phagocytic ability and MHC-II surface expression in most treatments. This research contributes toward understanding the genomic basis of cellular response differences to environmental stressors, demonstrates heat-LPS interaction, and genetic line-specific expression differences.
Support: USDA-NIFA-AFRI grant and Hatch project 5358.

W769: Poultry 1
Marek's Disease Virus Integration Profiles in Resistant and Susceptible Lines

W770: Poultry 1
Understanding the Functional Consequence of Selectively Constrained Regions of the Chicken Genome
Lel Eory, Roslin Institute University of Edinburgh, Edinburgh, United Kingdom, Alan L. Archibald, The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom, Paul Flicek, EMBL-European Bioinformatics Institute, Cambridge, United Kingdom and David W. Burt, Roslin Institute Univ of Edinburgh, Edinburgh, United Kingdom
Access to new genome assemblies of birds made it possible to identify regions in bird genomes which are of functional relevance. Regions which show lower level of divergence relative to an assumed neutral standard can indicate the presence of protein-coding genes and exons, regulatory RNAs, transcription factor binding sites as well as enhancer and silencer sequences. Here we present case studies based on our high-resolution map of selectively constrained elements for the chicken genome and how we can potentially link these with functional information.
In our study we created whole genome multiple sequence alignment of 48 birds and using the software GERP++ we called selectively constrained elements. This constraint map was then compared with the current chicken genome annotation and was integrated with 20, tissue specific, RNA-seq dataset.
Currently, less than 4% of the chicken genome is annotated as functional (Ensembl 75) and code for protein-coding and non-coding genes. In comparison we predicted 1.5 million selectively constrained elements, covering approximately 15% of the genome. Relative to the 18,000 transcripts and 17,000 genes present in Ensembl, we predicted over 200,000 transcripts and 50,000 genes based on our RNA-seq study. Several thousands of the new genes are predicted to be protein coding, based on their phylogenetic signatures, while the rest is potentially non-coding and may play a part in regulation. Our study clearly indicates that the functional complexity of the chicken genome is at present poorly understood, but potentially is on par with the mammalian genomes.

W771: Poultry 1
Transcriptome Response of Small Intestine of Village Chickens from Two Agro-Ecological Zones of South Africa Naturally Infected with *Ascaridia galli* parasites
Petunia D. Malatji[1], Farai C. Muchadeyi[1] and Este vanMarle-koster[2], (1)Agricultural Research Council, Pretoria, South Africa, (2)University of pretoria, Pretoria, South Africa
Nematodes of the genus Ascaridia are known to infect many species of birds and cause fatal diseases. *Ascaridia galli* damages the intestinal mucosa of chickens leading to blood loss and secondary infection. Very little information is available on the genetic resistance to gastrointestinal parasites particularly in village chickens populations. This study investigates the gene expression profiles in chickens from two different agro-ecological zones of South Africa that were naturally infected with *A. galli* parasites. Cox1 gene of *A. galli* parasities revealed six diverse haplotypes with no genetic differentiation between *A. galli* from Limpopo and KZN provinces. The naturally infected *A. galli* intestines displayed hypertrophy of the intestinal villi and necrosis of the crypt of Lieberkühn gland. Total RNA from small intestines of infected and non-infected village chickens was sequenced using Illumina HiSeq 2500 to generate between 3,908,924 and 3,994,946 reads. An average of 83.50% of trimmed reads were mapped to *gallus.galgal4.74* genome using Tophat program. Of 16383 DEGs detected between any two-way comparisons using Cuffdiff, 3057 DEGs were up/down regulated in Limpopo chickens whereas 1417 DEGs were up/down regulated in KZN chickens. A total of 426 DEGs were shared between the comparisons. Gene ontology analysis of DEGs revealed an enrichment of immune response, defense response, inflammatory response and cell signaling genes. T cell receptor signaling pathway was among the most significantly impacted pathways in both comparisons. The functional annotation of DEGs presents an opportunity to understand the molecular network underlying the genetic resistance of village chickens to *A. galli* infections.

W772: Poultry 2
MeSH Annotation of the Chicken Genome
Gota Morota, University of Nebraska-Lincoln, Lincoln, NE, Tim Beissinger, USDA-ARS, University of Missouri, Columbia, MO and Francisco Peñagaricano, University of Florida, Gainesville, FL
Biomedical vocabularies and ontologies aid in recapitulating biological knowledge. Annotation of gene products is mainly accelerated by Gene Ontology (GO) and more recently by Medical Subject Headings (MeSH). MeSH is the National Library of Medicine's controlled vocabulary and it is making inroads into annotation enrichment analysis. The availability of MeSH annotation in farm animals poses both new challenges and opportunities for downstream analysis in functional genomic studies. Here we investigated well-curated MeSH and GO terms to characterize the lists of selected chicken genes available at public repositories: differentially expressed genes from RNA-seq and genes subjected to sweeps and/or epistatic selection. The comparison of MeSH with GO overrepresentation analyses suggested not only that MeSH supports findings obtained from GO but also that MeSH is able to further enrich the representation of biological knowledge. Based on the hierarchical structures of MeSH and GO, respectively, we computed the semantic similarity between vocabularies as well as the similarity between selected genes. The respective hierarchies of MeSH and GO yielded the similarity levels between significant terms and the annotation of each yielded the measures

of gene similarity. Special attention is paid to a pair of genes showing high similarity in both MeSH and GO-based measures or in only one of them. We argue that the use of MeSH in conjunction with GO will be instrumental in facilitating the biological interpretation when a set of selected genes is identified.

W773: Poultry 2
TBD by Dr. Song

W774: Poultry 2
Intestinal Nutrient Transporter Expression in Embryonic and Posthatch Turkeys
Eric A. Wong[1], Melodie Weintraut[1], Sungwon Kim[2] and Rami A. Dalloul[1], (1)Virginia Tech, Blacksburg, VA, (2)The Roslin Institute R(D)SVS, Midlothian, United Kingdom
The absorption of nutrients in the small intestine is mediated by a variety of transporter proteins, which have not been as well characterized in turkeys as in chickens. The objective of this study was to profile the mRNA expression of aminopeptidase N (APN), 7 amino acid/peptide transporters (ASCT1, $b^{0,+}$AT, CAT1, EAAT3, LAT1, y^+LAT2, PepT1) and 3 monosaccharide transporters (GLUT2, GLUT5, SGLT1) in the small intestine of male and female turkeys pre-hatch (embryonic days 21 and 24 and day of hatch) and posthatch (day of hatch, days 7, 14, 21 and 28). Real-time PCR was used to determine mRNA abundance and data were analyzed by ANOVA using JMP Pro 11.0. APN, $b^{0,+}$AT, PepT1, y^+LAT2, GLUT5 and SGLT1 showed increased expression from E21 and E24 to DOH. Posthatch, all genes except GLUT2 and SGLT1 were expressed greater in females than males. GLUT2 was expressed the same in males as females and SGLT1 was expressed greater in males than females. The basolateral membrane transporters (ASCT1, CAT1, LAT1, y+LAT1, y+LAT2) were expressed greater during early development then decreased with age; while the brush border membrane transporters (EAAT3, GLUT5 and SGLT1) showed increased expression later in development. Turkeys showed high mRNA abundance of the anionic amino acid transporter EAAT3, which was 6-fold greater in the ileum of turkeys than chickens at D14. These results can be utilized to not only better formulate turkey diets to accommodate increased glutamate transport, but to also optimize nutrition for both sexes.

W775: Poultry 2
The Domestichick Project: from 57K to Whole Genome Sequence Data.
Michèle Tixier-Boichard[1], Mahendra Mariadassou[2], Pierre Nicolas[2], Mariangela Marca[2], Diane Esquerre[3] and Alain Vignal[4], (1)INRA, UMR GABI, Jouy-en-Josas, France, (2)INRA UR MAIAGE, Jouy-en-Josas, France, (3)Plateforme Genomique, Castanet tolosan, France, (4)INRA UMR GenPhySe, Castanet Tolosan, France
The red junglefowl (*Gallus gallus*) is considered as the main ancestor of the domestic chicken, but the genus *Gallus* includes three other species : *Gallus sonneratii* (grey junglefowl), *Gallus varius* (green junglefowl) and *Gallus lafayetii* (Sri Lanka junglefowl). Deep genome sequencing (25-30X) has been undertaken for these species in order to analyze with a great accuracy the genetic diversity within the genus *Gallus* and to better understand the genetic make-up of domestic chickens.
The individuals to be sequenced have been chosen according to their geographic origin (Vietnam, Thailand, Japan, Taiwan for wild *Gallus*; Europe, Africa, Asia, South America for domestic chickens) and on the basis of the genotypes obtained on the 57K SNP Illumina iSelect chicken array, in order to best represent the diversity of the sample. In total 18 wild individuals and 18 domestic chickens, 1 per breed, mostly females, were sequenced in paired-end.
The total number of SNPs detected as compared to the reference genome GalGal4 reached 42 millions. Some regions exhibited a strong differentiation between *G. sonneratii* and *G. gallus* (on chromosomes 7 and 28 for instance) whereas others exhibited a very low variability, suggesting a strong selective pressure (on chromosomes 5 and 18 for instance). The analysis of heterozygosity on the Z chromosome provides new insight into the pseudo-autosomal region. A HMM model applied to all data showed that the contribution of *G. sonneratii* to the genetic make-up of chickens could vary from 0.14 to 1.22% according to the breed.

W776: Poultry 2
Differential Expression Profiles of miRNAs Induced by Vaccination Followed by Marek's Disease Virus Challenge at Cytolytic Stage in Chickens Resistant or Susceptible to Marek's Disease
Huanmin Zhang[1], Shuang Chang[2], Qingmei Xie[3], Yanghua He[4], Kunzhe Dong[1], Catherine W. Ernst[5] and Jiuzhou Song[4], (1)USDA ARS Avian Disease & Oncology Laboratory, East Lansing, MI, (2)Shandong Agricultrural University, Tai'an, Shandong, China, (3)South China Agricultural University, Guangzhou, China, (4)University of Maryland, College Park, MD, (5)Department of Animal Science, Michigan State University, East Lansing, MI
Mounting evidence shows microRNAs (miRNAs) directly regulate gene expression post-transcriptionally through base-pairing with regions in the 3'-untranslated sequences of target gene mRNAs, which results in dysregulation of gene expression/translation and subsequently modulates cellular processes. We reported earlier that Marek's disease (MD) virus (MDV) induced significantly differentiated expression of 44 and 71 miRNAs in the line 6_3 MD resistant and line 7_2 susceptible chickens, respectively. We then reported 58 and 17 differentially expressed miRNAs induced by HVT, 17 and 57 miRNAs induced by CVI988/Rispens in the line 6_3 and 7_2 chickens, respectively. This study was designed to identify miRNAs differentially expressed in spleen in response to vaccination followed by MDV challenge at cytolytic stage by small RNA deep sequencing. Twenty four and 36 differentially expressed miRNAs were identified in response to HVT followed by MDV challenge, while Rispens and MDV induced 6 and 31 differentially expressed miRNAs in lines 6_3 and 7_2, respectively, in contrast to the control counterparts (Log_2 fold change ≥ 2.0). Hundreds up to thousands of target genes were predicted for the differentially expressed miRNAs. Over thirty pathways were likely involved with the lists of target genes of the differentially expressed miRNAs, which included TGF-beta signaling, MARK signaling, and Wnt signaling pathways *etc*., suggesting that the differentially expressed miRNAs are likely to play important roles in immunity, vaccinal protection, and suppression of tumorigenesis against MD through complicated networks. It is anticipated that further studies in extension of this project should lead to better understanding on how microRNAs mediate vaccine protective efficacy in chickens.

W777: Poultry 2

Genome-Wide Gene Express

Fiona M. McCarthy[1], Amanda M. Cooksey[1], Cathy R. Gresham[2], Ken Pendarvis[1] and Shane Burgess[1], (1)University of Arizona, Tucson, AZ, (2)Institute for Genomics, Biocomputing & Biotechnology, Mississippi State, MS

Recent projects aimed at identifying functional elements within key animal genomes and to sequence representatives of all avian clades highlight the importance of large-scale, experimental-based annotation of genomes. We have experimentally annotated the chicken genome by measuring mRNA, protein and ncRNA expression from multiple Red Jungle Fowl tissues (male and female). In addition to correcting both NCBI and Ensembl gene models and identifying new miRNAs and lncRNAs, we also quantitatively compare mRNA and proteins expression across multiple tissues and use this information to improve gene product annotation. We also provide fundamental information about gene expression, including tissue-specific expression of both RNA and protein gene products across 15 adult tissues (for both male and female birds), and demonstrate how expression is linked to function. All data generated by this project is submitted to SRA and PRIDE archives and gene expression results are freely and publicly available for viewing or download at the Chickspress (http://geneatlas.arl.arizona.edu/) website.

W778: Poultry 2

Chasing the Genetics of Ascites in Broilers

Douglas D. Rhoads, Katy Tarrant, Shatovisha Dey, Khaloud Alzahrani and Nicholas B. Anthony, University of Arkansas, Fayetteville, AR

Our research group has been pursuing the genetics of ascites in broilers for many years. Ascites, a form of pulmonary arteriole hypertension, is a collection of adverse changes in a broiler chicken including high venous pressure, accumulation of fluids in the body cavity, and cardiac hypertrophy, resulting from the reduced ability for oxygen to be supplied to the organs. These events result in either processing plant condemnation or death from heart and lung failure. Losses in the poultry industry vary but have been estimated at $10-100 million/year. We have employed low and medium density SNP-based genome wide association studies (GWAS) in multiple generations of experimental lines selected based on ascites phenotype. Earlier GWAS analysis identified regions on chromosome 9, but further analyses have yet to demonstrate these regions as generally applicable in commercial lines. Multigenerational GWAS identified regions on chromosome 2 and Z. Further analyses reveal that only the GgaZ region is reliably associated with ascites in our experimental line. Review of our results suggest there are no major genes for ascites in our experimental lines, despite an ability to rapidly select for or against ascites phenotype and a relatively high estimated heritability. Alternatives for moving forward on dissecting ascites susceptibility will be discussed.

W779: Poultry 2

Temperature Effects on Gene Expression in Turkey Satellite Cells.

Kent M. Reed[1], Kristelle M. Mendoza[2], Cindy Coy[3], Gale M. Strasburg[4] and Sandra G. Velleman[3], (1)University of Minnesota, St. Paul, MN, (2)University of Minnesota, ST. Paul, MN, (3)The Ohio State University, Wooster, OH, (4)Michigan State University, East Lansing, MI

As precursors to skeletal muscle, satellite cells mediate posthatch muscle growth. The function of satellite cells is affected by temperature with long-term potential effects on muscle growth, accumulation of intramuscular fat and the incidence of muscle myopathies. Newly hatched turkey poults and chicks have immature thermoregulatory systems yet are often exposed to acute temperatures immediately posthatch. Poultry selected for growth have an inefficient thermoregulatory system and are more sensitive to temperature extremes. This study examined gene expression in cultured satellite cells under thermal challenge (cold and hot). Turkey p. major satellite cells isolated from the RBC2 (control) and F (growth selected) lines were plated and replicate cell cultures were placed at 33 38 or 43C for 48 or 72 hr and RNA was extracted from harvested cells for RNAseq analysis. Paired-end reads were obtained for 24 libraries. Differential expression analysis found greater Time and Temperature effects than Line effects. Understanding how both cold and hot temperatures affect satellite cell proliferation and differentiation can be used to identify and develop thermal management strategies to improve skeletal muscle growth.

W780: Proteomics

Identification of Soybean Proteins and Genes Differentially Regulated in Near Isogenic Lines Differing in Resistance to Aphid Infestation.

Gary Stacey, University of Missouri, Columbia, MO

Soybean aphid, is an important pest causing significant yield losses. The *Rag*2 locus confers resistance to soybean aphid biotypes 1 and 2. Transcriptomic and proteomic analyses were done over a 48 h period after aphid infestation using near isogenic lines (NILs) differing at the *Rag*2 locus. Comparing the *Rag*2 and/or *rag*2 lines identified 3445 proteins with 396 differentially regulated between the two lines, including proteins involved in cell wall metabolism, carbohydrate metabolism, and stress response. RNA-seq transcriptomic analysis identified 2361 genes significantly regulated between the resistant and susceptible lines. Genes up-regulated in the *Rag*2 line were annotated as involved in cell wall, secondary and hormone metabolism, as well as in stress, signaling and transcriptional responses. Genes down-regulated in the *Rag*2 line were annotated as involved in photosynthesis and carbon metabolism. Interestingly, two genes (unknown and mitochondrial protease) located within the defined *Rag*2 locus were expressed significantly higher in the resistant genotype. The expression of a putative NBS-LRR resistant gene within the *Rag*2 locus was not different between the two soybean lines, but a second NBL-LRR gene located just at the border of the defined *Rag*2 locus was. Therefore, this gene may be a candidate R gene controlling aphid resistance.

W781: Proteomics

The Next Frontier: Proteome Dynamics and Fungal Pathogenicity.

Ralph A. Dean, Center for Integrated Fungal Research, Raleigh, NC

Magnaporthe oryzae, the causative agent of rice blast disease, infects plant leaves via formation of an appressorium which facilitates penetration of the leaf surface. In an effort to better understand the physiological changes accompanying the earliest stages of infection-related development, a nano-LC MS/MS-based global proteomics examination of conidial germination and cAMP-induced appressoria formation resulted in the

identification of over 5000 proteins. In developing appressoria, changes in cell wall modifiying, transport, extracellular and plasma membrane localized proteins were observed. A comparison of proteome and transcriptome data revealed little correlation between protein and transcript regulation. To better define the role of protein phosphorylation, study of the phosphoproteome was undertaken. A total of 2924 class I phosphosites were identified from 1514 phosphoproteins. Network analysis incorporating regulation from transcriptomic, proteomic and phosphoproteomic data revealed new insights into the regulation of the metabolism of conidial storage reserves and phospholipids, autophagy, actin dynamics and cell wall metabolism during appressorium formation. In particular, protein phosphorylation appears to play a central role in the regulation of autophagic recycling and actin dynamics during appressorium formation. Changes in phosphorylation were observed in multiple components of the cell wall integrity pathway as well as several transcription factors. Functional analysis of MGG_05709 provided further evidence for the role of protein phosphorylation in regulation of glycerol metabolism and the metabolic reprogramming characteristic of appressorium formation. Ubiquitination, plays a key role in the infection process. Current work is focused on examining global changes in phosphorylation-dependent ubiquitionation.

W782: Proteomics

A Proteomic Strategy to Discover the Composition, Localization, and Dynamics of Endogenous Protein Complexes in Arabidopsis Leaves

Dan Szymanski, Purdue University, W. Lafayette, IN

A given eukaryotic cell contains thousands of protein complexes that enable the cell to generate and respond to signals, fine tune and integrate metabolic activities, and carry out complicated mechanical tasks. Therefore broad knowledge about protein complex composition and dynamics is needed to analyze systems-level behaviors. To date, affinity purification-mass spectrometry and yeast-two-hybrid are the most widely used methods for high throughput analysis of protein complexes. However, these methods are not suitable for many plant species that are refractory to transformation or genome-wide cloning of open reading frames. To overcome these problems, we recently developed a new method that combines size exclusion chromatography with quantitative MS to analyze thousands of proteins leading to the discovery of hundreds of novel protein complexes (Aryal et al., 2014). We have recently expanded this technique in several important ways. First we have incorporated stable isotope labeling to anlayze how protein complexes rearrange in respose to metabolic stress. Second, we developed a new method to predict protein complex composition by combining SEC with an orthogonal ion exchange chromatography separation to generate abundance profiles of thousands of proteins using MS1 extracted ion chromatograms. These abundance profiles were subjected to clustering analysis to identify proteins that co-fractionate, and are putative subunits of native complexes. Last, our analysis has been extended to include membrane-associated proteins. Our update will include validations using both analyses of known protein complexes and Arabidopsis mutant in predicted complex subunits in order to identify true positives and false negatives.

W783: Proteomics

The Quest for Tolerant Varieties: The use of Proteomics to Understand Stress and Identify Variety Specific Alleles.

Sebastien C Carpentier[1], Ewaut Kissel[2], Jelle Van Wesemael[2] and Nadia Campos[2], (1)SYBIOMA, Leuven, Belgium, (2)KU Leuven, Leuven, Belgium

The primary objective of crop breeding is to improve yield and/or harvest quality while minimizing inputs. There is a need for improvement of crops and new tools to mine biodiversity for new promising alleles. While significant progress has been made in molecular and genetic analysis of model plants, the exploration of crop biodiversity and the correlation of cellular responses that can be analysed by omics to stress tolerance at the plant level is currently a challenge. New developments offered by proteomics techniques will contribute to this improvement.

We at KU Leuven host the Bioversity International collection of banana (more than 1500 accessions). We believe that a profound characterization of the available diversity enables to discover tolerance related alleles. Drought is a priority in banana breeding and drought tolerance is a complex quantitative trait, controlled by many genes. Due to the sterile nature of edible banana varieties traditional QTL mapping is not straightforward.

In 2011, we already proved that a gel based proteomics approach was able to select allele specific proteins involved in drought (Carpentier et al. 2011). Now we present a gel free based workflow to visualize the peptides that bear amino acid polymorfism(s) and are able to select those allele specific peptides that are correlated to the trait of interest via integrated multivariate analysis. Allele specific peptides are identified via clustering of MSMS spectra, de novo identification or searching against species specific mRNA databeses. Once performing high-throughput shotgun proteomics, annotation becomes quite challenging in a non-model crop. Therefore we additionally propose a cytoscape based workflow to cluster related proteins based on their GO annotations.

W784: Proteomics

Identification of Defense-related Proteins in the Root Necrotic Mutant *rn1* in Soybean

Madan K. Bhattacharyya, Priyanka Das and Narinder Pal, Iowa State University, Ames, IA

The root necrotic mutant *rn1* in soybean is caused presumably by insertion of the endogenous CACTA-element, *Tgm9*. It has been shown that excision of *Tgm9* from the *W4* locus resulted in a high frequency of mutation that negatively regulates the cell death process in *rn1* mutants. *Tgm9* in T322 genotype, generated from an Asgrow line, is highly active and we have recently demonstrated that the element transposes to all soybean chromosomes with a preference for genic regions. The *rn1* mutant was identified from *Tgm9*-induced mutant population and multiple mutations mapped to *rn1* locus and are allelic. Earlier it was shown that the mutant gains tolerance to *Phytophthora sojae*, the oomycete root pathogen. In this study we investigated the proteomes of necrotic and healthy roots identified from segregants of soybean plants heterozygous for *Rn1*. Proteins from roots of 5 to 7 day-old seedlings exhibiting either root necrosis or no root necrosis were isolated in the presence of a cocktail of phosphatase inhibitors. Protein samples were prepared individually from three biological replications for both (i) necrotic and (ii) non-necrotic roots and labeled with six iTRAQ reagents with molecular weights varying from 126 to 131. Six labeled protein samples (two treatments X three replications) were mixed prior to conducting LC-ESI-MS/MS. The proteins that accumulated in necrotic roots include several types of pathogenesis related proteins, leucine-rich repeat containing protein, trypsin and protease inhibitors, enzymes involved in antimicrobial compound biosynthesis, and glutathione S-transferases. Classification of phosphoproteins based on biological processes revealed that the

identified phosphoproteins in necrotic roots are involved in signal transduction; whereas, in healthy roots are involved in carbohydrate metabolic (80%) and transport (20%) processes.

W785: Proteomics

Leaf Proteome Profiling and Their Interactions To Determine Disease Resistance in Grape

Remy Babich, University of Maine, Orono, ME and Ramesh Katam, Florida A&M University, Tallahassee, FL

The mechanism of solar radiation transformed in the grape berry as a sink through grape leaves is the most important factor affecting the quantity and quality of the fruit yield. Florida Hybrid bunch (FH) grape is widely cultivated in US southeast region and is popularly known for wine and table grape. However, most commercial cultivars are highly susceptible to anthracnose and to various foliar diseases. Our goal was to determine molecular mechanisms involved in the grape leaf during the berry development and anthracnose disease tolerance. The specific objective of this study was to determine genetic diversity of leaf protein composition among different cultivars of Florida hybrid bunch grape during their berry development and better understand their role in disease tolerance characteristics. Leaf proteins from various FH cultivars were resolved on 2DE and characterized using Mass Spectrometry and the *Vitis* database. Comparative analysis of eight FH cultivars revealed 56 differentially expressed proteins mainly belonging to pathogen response, photosynthesis, and metabolism. Tolerant cultivars showed a greater overall abundance of proteins compared to susceptible cultivars. Interaction studies between photosynthetic related proteins suggest chlorophyll a/b binding proteins and photosystem 1 light harvesting proteins are more dominant in tolerant cultivars. Enzyme activity assay and real time PCR studies of selected genes and enzymes suggest a positive correlation with protein expression among cultivars. Protein expression, in particular photosynthesis related, in grape leaves is attenuated during berry development and ripening. Expression of antioxidant proteins such as superoxide dismutase and peroxidase declined in leaf at the early ripening stage suggesting a positive relationship connecting the source and sink organs of the FH bunch grapes.

W786: QTL Cloning

Map-based Cloning Reveals the Origin of *Fhb1* Gene in Wheat

Nidhi Rawat[1], Michael Pumphrey[2], Eduard Akhunov[1], James A Anderson[3] and Bikram S. Gill[1], (1)Kansas State University, Manhattan, KS, (2)Washington State University, Pullman, WA, (3)University of Minnesota, St. Paul, MN

Fhb1 (syn Qfhs.ndsu-3BS) is the most consistently reported QTL against the devastating Fusarium Head blight disease in wheat and has been reported to provide up to 60 % resistance in various mapping studies. *Fhb1* provides high level of type-2 resistance and is found in a few Chinese landraces only, from where it has been deployed worldwide. *Fhb1* from Chinese cultivar Sumai 3 has been cloned and validated in our Laboratory. Cloning of *Fhb1* opens up new avenues in understanding the origin and the mechanism of its action against FHB. Based on phenotyping and haplotyping, the origin of *Fhb1* is known to be restricted to a few Chinese landraces. Using sequence information obtained from the cloned gene, we studied the origin of *Fhb1* in wheat and and related species. We sequenced *Fhb1* gene from a large set of diploid A- genome, S- genome and D- genome accessions. Additionally, we sequenced it from a core set of tetraploid wheat species available with Wheat Genetics Resource Center at Kansas State University including *T. turgidum* and *T. timopheevii*. The findings of the association studies of *Fhb1* on wild and related germplasm of wheat research will be presented.

W787: QTL Cloning

Genetic and Physical Mapping of the Earliness *per se* Locus *EpsA^m1* in *Triticum monococcum* Identifies *EARLY FLOWERING 3* (*ELF3*) as a Candidate Gene

Maria Alejandra Alvarez[1], Gabriela Tranquilli[2], Silvina Lewis[2], Nestor Kippes[1] and Jorge Dubcovsky[1], (1)University of California, Davis, CA, (2)Instituto de Recursos Biológicos - INTA Castelar, Buenos Aires, Argentina

Wheat cultivars exposed to optimal photoperiod and vernalization treatments still exhibit differences in heading time. This variation, known as earliness *per se* (*Eps*), is important for the fine-tuning of flowering and the adaptation to different environments. We previously identified the *Eps-A^m1* locus from *Triticum monococcum* and showed that the allele from cultivated accession DV92 significantly delays flowering and increases the number of spikelets per spike relative to the allele from wild accession G3116. We expanded a high-density genetic map and physical map of the *Eps-A^m1* region and identified the wheat ortholog of circadian clock regulator *EARLY FLOWERING 3* (*ELF3*) as one of the candidate genes. No differences were found in *ELF3* transcript levels between NILs carrying the DV92 and G3116 *Eps-A^m1* alleles, but the encoded ELF3 proteins differed in four amino acids. These differences were associated with altered transcription profiles of *PIF-like, PPD1* and *FT1*, which are known downstream targets of ELF3. Tetraploid wheat lines with combined truncation mutations in the A- and B-genome copies of *ELF3* flowered earlier and had less spikelets per spike than the wild type control under SD and LD conditions. Both effects were stronger in the photoperiod sensitive than the photoperiod insensitive background, indicating an epistatic interaction between PPD1 and ELF3. By contrast, the introgression of the *T. monococcum* chromosome segment carrying *Eps-A^m1-l* allele from DV92 into durum wheat delayed flowering and increased the number of spikelets per spike in the field, providing a novel allele to modulate flowering time and spike development in wheat.

W788: QTL Cloning

Identification of Genes Controlling Earliness *per se* and Short Day Photoperiod Response in Bread Wheat

Simon Griffiths and Meluleki Zikhali, John Innes Centre, Norwich, United Kingdom

Major genes controlling photoperiod and vernalization sensitivity have been cloned. A knowledge gap exists in the identification of genes controlling Earliness *per se* (*Eps*) and short day specific (SDS) genes. We describe the positional cloning of an *Eps* locus on wheat 1DL named *Eps-D1*. *Eps-D1* is due to a deletion including several genes in the sub-telomeric region. Using near isogenic lines (NILs), and key recombinants, we identified *TaELF3-D1* within *Eps-D1* a homologue of the *Arabidopsis thaliana* circadian clock gene *ELF3* as the most likely candidate for *Eps-D1*. Gene expression, using NILs, showed altered *TaGIGANTEA* (*TaGI*) expression consistent with an ELF3 mutation. We also describe the identification of two short day specific QTLs on wheat 1BS and 1DL using doubled haploid populations grown under short and long days. Using genomics, bioinformatics and synteny with *Brachypodium distachyon*, we identified *TaTOE1-B1* and *TaFT3-B1* the homologues of *ZmTOE1* and *HvFT3* as the more likely candidates for the QTLs on 1BS and 1DL respectively. Gene expression analysis suggests that *TaTOE1* homoeologues

are repressors of *FT1*, expression of *TaTOE1* was higher in the juvenile plant, and under short days a result which was consistent with the role of *TOE1* in *Arabidopsis thaliana* suggesting a similar function in both species. *TaFT3* homoeologues were also expressed at higher levels under short days in barley. These genes will give options for breeding by design using the KASP markers we designed and will also enable further academic research to unravel the complex flowering time regulation pathways in temperate cereals.

W789: QTL Cloning

Speeding up QTL Cloning in Maize: Power and Prospects of the MAGIC Maize Population

Matteo Dell'Acqua[1], Daniel M Gatti[2], Elisabetta Frascaroli[3], Gary A Churchill[2], Dirk Inzé[4], Michele Morgante[5] and Mario Enrico Pè[1], (1)Scuola Superiore Sant'Anna di Pisa, Pisa, Italy, (2)The Jackson Laboratory, Bar Harbor, ME, (3)University of Bologna, Bologna, Italy, (4)VIB Department of Plant Systems Biology, Gent, Belgium, (5)Università di Udine, Udine, Italy

The first step to QTL cloning is the discovery of genetic loci linked to phenotypes of interest. Increasing the power, definition and speed by which such genetic loci are pinpointed on the genome is crucial for the success of this challenging task. New breeding designs emerged to address this need by increasing diversity and recombination events in QTL mapping panels. We discuss the MAGIC maize, a 1,636 recombinant inbred lines mapping panel we developed by intercrossing eight diverse *Zea mays* inbred lines in a balanced scheme. Each MAGIC recombinant inbred line is a unique mosaic of the diversity of the original founders, the population altogether collecting more than a hundred thousand recombination events. The increased diversity in elevated minor allele frequency permits the MAGIC maize to bridge the flexibility of association mapping approaches to the power of linkage mapping, speeding up the discovery of candidates for cloning. The MAGIC maize leverages transcriptomics data and full sequencing of the founder lines to reach single-gene definition. We genotyped and phenotyped a subset of the MAGIC maize and used it to show the general features of the population, identifying candidate genes for complex phenotypes. Mapping power simulations showed that the MAGIC maize may allow efficient QTL mapping even when few hundred lines are phenotyped, favoring its use in controlled conditions experiments. We discuss the perspectives that the MAGIC maize discloses in maize QTL mapping and cloning, and the further steps to exploit this resource, including the production of recombinant intercrosses for the study of heterosis.

W790: QTL Cloning

Cloning *qAG-9-2*, a Major QTL for Anaerobic Germination Tolerance in Rice

Endang M. Septiningsih[1], Tobias Kretzschmar[2], Margaret Anne F. Pelayo[2], Kurniawan R. Trijatmiko[2], Lourd Franz M. Gabunada[3], Rejbana Alam[4], Rosario Jimenez[2], Merlyn S. Mendioro[3], Inez Slamet-loedin[2], Nese Sreenivasulu[2], Julia Bailey-Serres[4], Abdelbagi M. Ismail[2] and David J. Mackill[5], (1)Texas A&M University, College Station, TX, (2)International Rice Research Institute, Los Banos, Philippines, (3)University Philippines Los Banos, Los Banos, Philippines, (4)University of California, Riverside, Riverside, CA, (5)University of California-Davis, Davis, CA

Tolerance to flooding during germination conditions or anaerobic germination (AG) is an important trait for direct-seeded rice (DSR), especially in the tropical regions. This trait allows rice seeds to germinate and survive under flooding stress due to heavy rains right after sowing. Weed invasion, a major problem for DSR, can also be significantly controlled by rice varieties having AG tolerance. Here we reported further investigation of a major QTL for AG tolerance in rice, *qAG-9-2*, derived from Khao Hlan On, a landrace from Myanmar. Near isogenic lines (NILs) carrying the QTL were developed and subsequent fine mapping delineated the region to a 50kb-DNA fragment. Through over expression and loss-of-function mutant studies, a trehalose-6-phosphate phosphatase gene, *OsTPP7*, was confirmed as the causal gene underlying *qAG-9-2*. This gene is involved in trehalose-6-phosphate (T6P) metabolism, central to an energy sensor that determines anabolism or catabolism depending on local sugar availability. Under AG stress, *OsTPP7* escalates sugars and amino acid pools, promotes anaerobic metabolism and increases transcripts associated with elongation growth. It is expressed in germinating embryos, coleoptiles and young roots, all sink tissues that depend on reserve carbohydrates for proliferation. OsTPP7 activity may increase sink strength in proliferating heterotrophic tissues by indicating low sugar availability through increase T6P turnover, thus enhancing starch mobilization to drive growth kinetics of the germinating embryo and elongating coleoptile, which consequently enhances AG tolerance. Marker-assisted breeding to introgress this gene into several Sub1 and non-Sub1 lines has been conducted.

W791: QTL Cloning

Tapping Native Diversity: Characterizing the Genetic Architecture of Complex Traits in Maize Landraces

Jorge Alberto Romero Navarro, Cornell University, Ithaca, NY

Landraces represent a good source for useful alleles to improve yield, especially under conditions of biotic and abiotic stress. Until recently, the identification and deployment of such alleles has been slowed by the lack of phenotypic association with genetic markers and their linkage with numerous undesirable alleles. In addition, because of historic recombination, landraces offer the potential for resolving at high resolution regions associated with quantitative variation.

Here, we present the results for performing genome wide association for flowering time and plant height in a panel of 3,500 landraces from Latin America. Using Genotyping by Sequencing, we generated around 1 million SNPs for the accessions sampled. We performed phenotypic evaluation on 23 locations across Mexico for 2 years. With these data, we have identified association between flowering time and genes known to be part of its genetic architecture. Furthermore, we report association at hundreds of novel genes, as well as association at large structural variation like inversions and centromeres. For plant height, we show significant association at various genic regions. For both traits, we explore the predictive ability of the most significant markers and contrast the differences in predictive ability between both traits according to their genetic architecture.

This study provides better knowledge about the genetic architecture of two agronomically important traits, highlights the effect of genome structure variation on quantitative traits, and reveals the potential for uncovering useful alleles from very diverse populations

W792: QTL Cloning

A QTL Cloning Pipeline in Wheat using MAGIC

Klara Verbyla, CSIRO, Digital Productivity and Agriculture Flagships, Canberra, Australia

CSIRO has developed two wheat MAGIC populations that are being utilized to improve our understanding of the underlying genetic basis of a range of quantitative traits. The population with four founders has been extensively deployed for more than 5 years, while the population with eight founders has become more heavily utilized in recent years. This has resulted in extensive QTL data for a range of complex traits. As a result a novel QTL validation and cloning pipeline has been implemented. The pipeline exploits the increased allelic diversity, power and mapping resolution afforded by large multi-parent mapping populations, whilst reducing complexity by using multi-allelic contrasts at the targeted QTL region. The genetic diversity within the populations also means that the resource is useful for a wide range of traits and genetic backgrounds. I will provide an overview of the pipeline and an application of the approach.

W793: Recombination - mechanisms

DNA Rearrangement Mapping Reveals Surprising Patterns of Genomic Instability in Organelles

Samuel Tremblay-Belzile, Etienne Lepage, Éric Zampini and Normand Brisson, Université de Montréal, Montréal, QC, Canada

Genomic instability in organelles is linked to several phenotypes, including cytoplasmic male sterility in plants and some cancers and neurodegenerative disorders in mammals. To obtain a global portrait of genomic rearrangements, we developed an approach for the analysis of next-generation sequencing data from *Arabidopsis thaliana*. Our study reveals an unexpected abundance of short-range rearrangements in all samples, with U-turn-like rearrangements representing approximately 25% of total rearrangements in wild-type plastids. Comparing the patterns of genomic instability in different mutant lines suggest that U-turn-like events correlate with replication stress, and are inhibited by Whirly and RecA proteins. A 60-fold increase in U-turn-like rearrangements in the *why1why3reca1* mutant line correlates with a white-variegation phenotype with altered leaf shape. We therefore propose that U-turns act as a RecA-independent mechanism to restart stalled forks, and are favored by the presence of nearby microhomologies in inverted orientation. They are also observed in large proportions in mitochondria, both in *Arabidopsis thaliana* and humans, raising the question of whether they might be involved in other phenotypes or diseases involving organelles.

W794: Recombination - mechanisms

Overview of Actual and Ancestrale Recombination in Relationship with the Sequence in Wheat : Focus on 3B Chromosome

Benoît Darrier[1], Alain Loussert[2], Hélène Rimbert[1], Julien Navarro[2], Frederic Choulet[2], Philippe Leroy[2], Lise Pingault[2], Aurelie Evrard[2], François Balfourier[1] and Pierre Sourdille[1], (1)INRA GDEC, Clermont-Ferrand, France, (2)INRA UMR 1095 GDEC, Clermont-Ferrand, France

Meiotic recombination (crossovers) is a mechanism that largely contributes to shape the genome structure in all species. However, in bread wheat (*Triticum aestivum* L.), recombination occurs almost exclusively in distal telomeric regions. Thus, many genes contained in pericentromeric regions which represent 80% of the chromosomes are not admixed during meiosis. We exploited the chromosome-survey data from IWGSC to study at the genome scale 596 CO events located on 445 contigs. Among these, more than 250 CO mapped on 25 scaffolds of the 3B pseudomolecule on features smaller than 50kb. This number was increased to reach ~500 COs on 3B which were correlated to sequence features (epigenetic landmarks, genes, transposable elements, specific motifs, RNAseq data of meiosis) available for chromosome 3B. We thus showed that COs mainly occur in the vicinity of genes. Moreover, we used two collections of 90 lines corresponding to Asian and European genetic pools to study the variation of linkage disequilibrium (LD) as well as the ancestral recombination at a finer scale in the regions covering our previously detected hot spots. Analysis revealed high difference in LD structure between the two populations underlying the complex history of each genetic pool. However historical mapping of recombination events using the same SNPs showed a common location of recombination breakpoints with variable intensity. Impact of these results on the improvement of recombination in wheat will be discussed.

W795: Recombination - mechanisms

The Stability and Consequences of Recombination in Maize

Eli Rodgers-Melnick, Institute for Genomic Diversity, Cornell University, Ithaca, NY, Daniel L. Vera, Florida State University, Tallahassee, FL, Peter Bradbury, Cornell University/USDA-ARS, Ithaca, NY, Hank W. Bass, Department of Biological Science, Florida State University, Tallahassee, FL and Edward S. Buckler, USDA-ARS-Cornell University, Ithaca, NY

Meiotic recombination varies by over 3 orders of magnitude within the genomes of many eukaryotic organisms, including crop species such as maize. Because plant breeders rely on meiotic cross-overs to introgress beneficial alleles and fine-map quantitative traits, the tremendous regional variation in cross-over rates fundamentally limits the practice of crop improvement. Population genetic theory further predicts that deleterious mutations should preferentially accumulate in low-recombination regions, particularly in historically outcrossing species such as maize. Here, we show that meiotic recombination is predictable across diverse maize crosses, with epigenetic indicators of open chromatin delineating the locations of recombination hotspots and sites of GC-biased gene conversion. We further demonstrate that the extant patterns of recombination are historically stable and tied to variation in the frequency of deleterious mutations. The ability of plant breeders to exploit recombination in purging segregating deleterious alleles will determine the efficacy of future crop improvement.

W796: Recombination - mechanisms

The Effect of Temperature on the Male and Female Recombination Landscape of Barley

Luke Ramsay, The James Hutton Institute, Dundee, United Kingdom

In barley (*Hordeum vulgare* L.) and potentially many other members of the Pooideae, up to a third of genes are largely inaccessible to conventional breeding programmes as meiotic crossovers are localised to the ends of the chromosomes. There is however evidence that environmental stress can affect crossover patterns so we utilised a genome-wide marker set for linkage analysis combined with cytological mapping of crossover events to examine the recombination landscape of plants grown at different temperatures. We found that barley shows heterochiasmy i.e. differences between female and male recombination frequencies at higher temperatures. In addition, we found that elevated temperature significantly changes patterns of recombination in male meiosis only, with a re-positioning of Class I crossovers determined by cytological mapping of HvMLH3 foci. We show that the length of synaptonemal complexes in male meiocytes also increases in response to temperature

The results demonstrate that the distribution of crossover events are malleable in barley and can be shifted to proximal regions by altering the growth temperature. The shift in recombination is the result of altering the distribution of Class I crossovers, but the higher recombination at elevated temperatures may not be the result of solely an increase in Class I events.
The implications of these results will be discussed in terms of the control of recombination landscape in barley as well as the potential applications to cereal breeding programmes.

W797: Recombination - mechanisms

Brassica Allotriploids Hybrids: New Way to Exceed Recombination Limits for COs Rate and Distribution

Alexandre Pelé[1], Gwenn Trotoux[1], Frederique Eber[1], Sylvie Nègre[1], Maryse Lodé[1], Marie Gilet[1], Olivier Coriton[1], Virginie Huteau[1], Matthieu Falque[2], Olivier C. Martin[2], Jérôme Morice[1], Mathieu Rousseau-Gueutin[1] and Anne-Marie Chevre[1], (1)INRA, Le Rheu, France, (2)INRA/CNRS/Univ Paris-Sud/AgroParisTech, Gif-sur-Yvette, France

Meiotic homologous recombination by crossovers (COs) is the main mechanism responsible for mixing genetic diversity in plant breeding. However, loci separations are limited due to the strict regulation of the rates and distribution of recombination events along the chromosomes. Indeed, rarely more than two COs occur between homologous chromosomes per meiosis and their distribution is not homogenous in all chromosomic regions, such as the pericentromeric regions that are free of recombination. Exceptions to this rule were observed and linked to different factors such as polyploidy, genetic background or sex meiosis. Accordingly, it was shown that in allotriploid hybrids (AAC, 2n=29), resulting from crosses between *Brassica napus* (AACC, 2n=38) and *B. rapa*(AA, 2n=20), COs get a boost along the A genome compared to diploid hybrids (AA, 2n=20). However, the impact on COs distribution as well as the effects of different genetic backgrounds and sex meiosis were still unclear. To that purpose, progenies deriving from 3 diploid and 3 triploid hybrids were used, enabling to study the distribution of 3000 COs per hybrid by the genotyping of 200 SNPs well distributed along A chromosomes (one SNP each 1.2 Mb). Compared to what was previously known, we showed that the presence of the haploid C genome modifies the COs distribution, by generating new recombining regions within A chromosomes. In addition, significant variations were observed between triploid hybrids depending on the genetic background and sex meiosis. These findings may enable to combine new loci by modifying interference and linkage disequilibrium.

W798: Recombination - mechanisms

Natural Modifiers of Recombination in Arabidopsis

Ian R Henderson, University of Cambridge, Cambridge, United Kingdom

During meiosis homologous chromosomes pair and undergo reciprocal crossover, which recombines linked genetic variation. Crossover frequency is under tight genetic control and eukaryotes typically experience one or a small number of exchanges per chromosome pair per meiosis, despite larger numbers of initiating DNA double strand breaks (DSBs). For example, Arabidopsis undergoes 100-200 DSBs that mature into ~10 crossovers per meiosis. However, variation in crossover frequency is also observed within and between species. Recombination can accelerate adaptation by combining independently arising mutations. Therefore, genes that modify crossover levels or patterns will profoundly influence selective responses throughout the genome. To further investigate this phenomenon we screened for natural variation that modifies meiotic crossover rate in the model plant Arabidopsis. Crossover-modifying polymorphisms can be defined as acting in either *cis* or *trans*, according to whether they control recombination on the same chromosome, or throughout the genome respectively. Segregation of linked, heterozygous T-DNAs expressing different colors of fluorescent protein in pollen or seed can be used to measure crossover frequency in Arabidopsis. We previously analysed crossover frequency in an F_2 population originating from a cross between the *420* sub-telomeric FTL and Catania-1 (Ct-1) and did not identify significant *trans* modifiers. To further screen for *trans* acting loci we generated a second Col-*420* x Landsberg *erecta* (Ler) F_2 population. We identified two major *trans* recombination QTLs (*rQTLs*) on chromosomes 1 and 4, with LOD values of 40.2 and 53.5 respectively. $rQTL1^{Ler}$ genotypes are associated with low recombination, with heterozygotes showing an intermediate crossover frequency. In contrast, $rQTL4^{Ler}$ behaves recessively and causes high recombination. In this talk I will present progress on identifying the causal polymorphisms underlying these recombination QTLs and mechanistic understanding of their role in crossover control.

W799: Resources and Programs for Undergraduate Education in Genomics

Intrductory Remarks

Scott Woody, UW-Madison, Madison, WI

Welcome, session goals, speaker introductions

W800: Resources and Programs for Undergraduate Education in Genomics

FPsc: A New, Plant-Based Model System for Integrated Education in Genetic and Genomic Sciences

Scott Woody, UW-Madison, Madison, WI

FPsc (“Fast Plants, self-compatible”) is a self-fertile analog of the self-incompatible Wisconsin Fast Plants (WFP) variety of rapid-cycling *Brassica rapa*. Like WFP, the FPsc variety progresses rapidly through the plant life cycle, flowering ~18 DAP and yielding mature progeny seeds 7-8 weeks after planting. Unlike the WFP variety, in which heterozygosity exists throughout the genome, FPsc is homozygous at >99% of all loci due to imposition of inbreeding (selfing with single seed descent) during our selective breeding program. We have exploited the properties of self-compatibility and homozygosity in FPsc to create an integrated suite of resources useful for hands-on education in genetic and genomic sciences.
Mutagenesis screens were used to identify a collection of both dominant and recessive FPsc mutant alleles whose phenotypes are engaging and clear. Concurrently, our students developed a complementary collection of PCR-based genetic marker assays that have been used to map and then sequence candidate FPsc loci to identify causative mutations. Genomic resources based on the FPsc model system will be described, including a *de novo* genome sequence assembly, RNA-Seq datasets obtained by study of several FPsc mutants, and Advanced Intercross RIL populations with proven utility to map QTL that condition quantitative traits in *B. rapa*.
This student-friendly suite of resources allows educators to realize the Holy Grail of genetics education: To enable students to connect observable phenotypic differences with the underlying, DNA sequence-based, genotypes.

W801: Resources and Programs for Undergraduate Education in Genomics

SEA-PHAGES: A Robust Integrated Course-Based Research Experience (iCRE) for Engaging in Authentic Research with Undergraduate Students

Welkin Pope, University of Pittsburgh, Pittsburgh, PA

The PCAST report in 2012 noted that the United States was not producing enough STEM majors to meet its future economic demand. One recommendation to ameliorate this issue was for universities to offer course-based research experiences (CREs) to engage early career scientists in authentic research in the classroom. To date, a number of CREs have been developed at institutions nationwide, the majority of these are local implementations (lCREs, with a very few integrated CREs (iCREs) disseminated across institutional boundaries. The Science Education Alliance – Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) (http://seaphages.org) project is an iCRE with over 90 member institutions nationwide (and one international member), and more than 8000 students engaged. SEA-PHAGES is a two-semester course that leads students through virus discovery and characterization at the bench in the fall, followed by in the spring by genome annotation and comparative genomics. The success of SEA-PHAGES lies in its central question which supports massively parallel investigation at all types of institutions. The program has yielded over 20 peer-reviewed scientific papers analyzing the collection of genomes, and has demonstrated educational gains through improved retention of students in STEM majors, elevated grades, and increased feelings of scientific efficacy and sense of community.

W802: Resources and Programs for Undergraduate Education in Genomics

Discovery Based Modules for Illustrating mRNA Splicing, Genome Variation and Detecting Genetically Engineered Organisms

James Burnette, University of California, Riverside, Riverside, CA

We have developed several modules to introduce key concepts in biology and genomics for the Dynamic Genome (DG) course offered at UC, Riverside. During this course based research experience (CURE) first-year students learn about genetic information transfer (Central Dogma) and genome variation through experimentation on plant DNA. Students will extract DNA, perform PCR and agarose gel electrophoresis as well as use bioinformatics to analyze DNA sequence data in the first half of the course. In the second half of the course, students complete an open-ended research project based on ongoing research at UCR. I will present four modules that are used in class. The first one introduces dilution and pipetting. The second module demonstrates the result of splicing mRNA and allows students the opportunity to learn PCR and basic bioinformatics. Student can also discover interspecific genome variation. The third module demonstrates intraspecific genome variation using the same techniques as the first module. The fourth module I will present teaches students the science behind genetic engineering giving them the facts so that they can make informed opinions of GMO plants and animals. Finally, I will demonstrate the electronic notebook we have developed for use in the teaching laboratory. I will briefly discuss how the protocols can be modified for short class periods found at high schools. The protocols and background materials for these modules will be made available on a website for participants to access.

W803: Resources and Programs for Undergraduate Education in Genomics

Getting on Track: DNA Barcoding Projects As a Vehicle to Engage Community College Students in Basic Life Sciences Research.

Alejandro Cortez, UC-Riverside, Riverside, CA

We have developed the “Sequencing to Success DNA Barcoding Challenge,” a module based on DNA barcoding that introduces students to the process of science from developing a question, experimentation, data analysis, and presentation of results. While similar to the module developed by the Dolan DNA Learning Center, we have targeted community college transfer students and high school teachers. These participants have limited time to spend in a research facility so we do the DNA analysis and poster preparation using interactive online platforms. The module begins with the question “Is the fish at the store properly labeled?” or “What plant is this?” Participants bring their sample to the lab. During their visit, students extract DNA from the sample, perform PCR using 16S (fish) or RBCL (plants) primers and analyze the results on agarose gels. During down time the students learn to use the electronic notebook (e-LN) platform and Google Drive where data will be exchanged and discussed. During the next week, the PCR products are sequenced and the results are shared electronically. Over the course of several weeks, students work with instructors in person or through Skype/Google-Hangouts to learn the DNA Subway for data analysis. Participants write an abstract that is reviewed and corrected by the instructors. Lastly, students prepare a poster for a symposium at UCR three to four weeks after the laboratory work. This program has served ninety-six community college students from Riverside and San Bernardino Counties over three years. Protocols will be made available at the presentation.

W804: Resources and Programs for Undergraduate Education in Genomics

Integration of RNA-Seq Data Analysis into Undergraduate Laboratory Teaching Modules

Ray A. Enke, James Madison University, Harrisonburg, VA

Undergraduate students learn about Next Generation Sequencing (NGS) platforms in courses, but often have difficulty understanding the impact of these techniques without hands-on experience analyzing actual NGS datasets. The NSF-supported ***Infrastructure & Training to Bring NGS Analysis Into Undergraduate Education*** **project** focuses on streamlining RNA-sequencing (RNA-seq) data analysis using a combination of bioinformatics and wet lab workflows tailored for implementation into diverse undergraduate laboratory classroom settings. These modular workflows can be applied to a variety of novel or publicly available RNA-seq datasets.

To create a course-based undergraduate research experience focusing on NGS that complements my research program, I developed an RNA-seq dataset to identify differentially expressed genes between embryonic day 8 (E8) and E18 chicken retinas. Computational and wet lab modules analyzing this RNA-seq dataset were implemented into an upper level undergraduate Advanced Molecular Biology course. Computational modules investigated gene ontology of the dataset to identify target candidate genes for follow up analysis as well as sequence retrieval, sequence annotation and gene-specific primer design. These modules were used in conjunction with wet lab activities constructing cDNA libraries from total RNA followed by quantitative reverse transcriptase PCR (qRT-PCR) using student-designed qPCR primers. Additional computer-based

modules were employed for analysis of qRT-PCR gene expression data. Lab modules and lesson plans for this and other courses are publicly available to educators via the "RNA-seq for the Next Generation" website developed by the Cold Spring Harbor Laboratory DNA Learning Center.

W805: Resources and Programs for Undergraduate Education in Genomics

What DNA Says About Our Human Family: Putting Students in the Tree of Human Evolution

Dave Micklos, DNA Learning Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Written in each person's DNA is an unbroken record of our shared ancestry as human beings – where we arose and how we moved and mingled to create the tapestry of cultures we now see around the world. Although people may look rather different on the outside, genetic analysis shows that all humans are closely related. DNA studies confirm that all people alive today are descendants of the first modern humans, who emerged in Africa about 200,000 years ago, and that we interbred with ancient Neanderthals and a mysterious group called the Denisovans. Genes also bear the marks of our mastery over other living things – the domestication of plants and animals. Learn about the DNA Learning Center's integrated biochemical and bioinformatics (B&B) workflow that uses students' own mitochondrial (mt) DNA sequences to discover principles of genetic diversity and human evolution. In class, students use a resin-based method to extract DNA from cheek cells, amplify a hypervariable region of the mt genome by PCR, and confirm results with gel electrophoresis. Student samples are then sent for DNA sequencing, and results are automatically uploaded within 48 hours into *Bioservers* and *DNA Subway* – intuitive workflows for sequence alignment and tree building. Students' own questions about their relatedness to classmates, individuals from world populations, and ancient hominids then drive their bioinformatics exploration of sequence data – culminating by placing themselves in their own tree of human evolution.

W806: Rice Functional Genomics

Nitrogen Use Efficiency: Transport Solution in Rice Variations

Chengcai Chu, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Asian cultivated rice (*Oryza sativa* L.) consists of two main subspecies, *indica* and *japonica*. *Indica* has much higher nitrate absorption activity over *japonica*, but its molecular mechanism remains elusive. Here we show that variation in a nitrate transporter gene, *NRT1.1B/OsNPF6.5*, may largely contribute to this nitrate use divergence. Phylogenetic analysis revealed *NRT1.1B* diverges between *indica* and *japonica*, and the *NRT1.1B-indica* variation not only enhances nitrate uptake and root-to-shoot transport, also up-regulates expression of nitrate responsive genes. The signature of *NRT1.1B-indica* subjected to artificial selection unveiled the possible origin of nitrate use divergence during rice domestication. Notably, field tests with either near-isogenic or transgenic lines confirmed that *japonica* variety carrying *NRT1.1B-indica* allele had a significant improvement of grain yield and nitrogen use efficiency (NUE). Our results demonstrate that variation in *NRT1.1B* largely explains nitrate use divergence between *indica* and *japonica*, and that *NRT1.1B-indica* has great potential for improving NUE of *japonica*.

Key words: Rice; *NRT1.1B*; nitrate; nitrate transporter; nitrogen use efficiency

W807: Rice Functional Genomics

A Powerful Tool for Efficient Recurrent Selection: Positively and Negatively Selectable Male-Sterile Rice

Junichi Tanaka[1], Maiko Akasaka[2], Yojiro Taniguchi[2], Masao Oshima[3], Kiyomi Abe[3] and Yutaka Tabei[3], (1)National Agriculture and Food Research Organization, Tsukuba, Japan, (2)NARO Institute of Crop Science (NICS), Tsukuba, Japan, (3)National Institute of Agrobiological Sciences (NIAS), Tsukuba, Japan

Recurrent selection is a powerful breeding method of maize: populations are improved by shuffling the many types and highly diverse genomes of maize through its allogamous mode of reproduction. To use the same selection method in an autogamous crop like rice, breeders are required to do several and continuous artificial crossings, furthermore recurrent selection is difficult to perform on a large scale. To circumvent these issues, we propose a novel method based on the development and application of a negatively and positively selectable, single-locus dominant transgenic male sterility. This method is will allow for an efficient recurrent selection in autogamous crop species (Tanaka 2010). The male sterility trait is used only during the breeding procedure and will be removed from the final products ('null segregant') since male-sterile individuals cannot self-pollinate to produce inbred lines. Currently at the NARO institute of crop science and NIAS, the development of transgenic rice for an efficient recurrent selection is in progress. We have developed anther-specific expression promoters from a comprehensive study using the expression profile database RiceXPro. The developed promoters were confirmed to be effective in producing dominant male-sterile individuals. Herbicide tolerance genes can be used as positively selectable markers. Fluorescent protein expressed under the gluteline promoter accumulate in the endosperm (Qu and Takaiwa 2006) and can be efficiently used as either positive or negative selectable, visible markers. We are now in the prototype development stage and will present our study and envisioned future applications.

W808: Rice Functional Genomics

Towards a Fully Indexed Mutant Population in the Model Rice, Kitaake, Suited for Forward and Reverse Genetics Research

Mawsheng Chern[1], Rashmi Jain[1], Guotian Li[1], Zhongchen Zhang[2] and Pamela Ronald[1], (1)UC Davis/JBEI, Davis, CA, (2)UC Davis, Davis, CA

We have generated a rice mutant population containing 7,000 independent (M1) lines using fast-neutron induced mutagenesis in the model rice Kitaake, a short-statured, short life-cycled, early flowering rice variety. In collaboration with the Joint Genome Institute, we aim to sequence and analyze 4,000 rice mutants in order to create a rice mutant population resource that reaches an estimated 90% coverage of total rice genes. Currently, we have sequenced over 1000 mutants. On average, each mutant contains over fifteen mutated genes. We have established a database, named KitBase, to integrate all information of this mutant population, including sequences, mutated genes, and available phenotypes. DNA sequence, gene or mutant ID, chromosome coordinates, and key words can be used to search KitBase.

Using this rice mutant population, we have conducted a forward genetic screen to identify mutants that suppress the rice immune response mediated by XA21, a pattern recognition receptor conferring robust resistance to *Xanthomonas oryzae pv oryzae* (*Xoo*). We have obtained 10 *suppressors of Xa21-mediated immunity* (*sxi*), after screening approximately 21,000 plants, and characterized 4 of the *sxi* mutants. We have used

a comparative genome hybridization approach based on rice whole-genome tiling arrays and a whole genome resequencing approach to identify candidate mutations in each mutant. The responsible mutation that cosegregates with the *sxi* phenotype was identified in the backcrossed segregating F2 progeny. We have complemented mutant *sxi1* with one of the candidate gene. This study demonstrates the power and usefulness of this rice mutant population.

W809: Rice Functional Genomics

GS + *de novo* GWAS in Tropical and Temperate Irrigated Rice Breeding Programs

Jennifer E. Spindel, Cornell University, Ithaca, NY

Genomic selection (GS) helps accelerate the rate of genetic gain in breeding by utilizing whole genome data to predict the breeding value of offspring. Genomic selection has been widely applied to maize and small grains breeding, but has only recently been applied to rice. Here, we discuss the results of GS cross-validation in two rice breeding programs: 1. the International Rice Research Institute (IRRI) irrigated rice breeding program for Southeast Asia (tropical rice), and 2. the National Institute for Agronomic Research in Uruguay (INIA) irrigated rice breeding program (temperate rice). Both breeding populations were genotyped using genotyping-by-sequencing (GBS) and genomic prediction performed for grain yield, plant height, and flowering time. In addition, prediction was performed in the INIA population for milling yield and grain chalkiness. RR-BLUP accuracies in both populations averaged ~0.3 for grain yield, however, by combining RR-BLUP with markers fit as fixed effects selected from the results of a genome-wide-association study (GWAS) on the RR-BLUP training data in the IRRI population we improved prediction accuracies up to 30% depending on the trait. We also show how making careful use of multi-environment data can improve prediction accuracies up to 10-fold over single environment models.

W810: Rice Functional Genomics

SNP-Seek: the Largest Database of Rice Natural Variations.

Nickolai Alexandrov, International Rice Research Institute, Los Baños, Laguna, Philippines

SNP-Seek is the largest database of rice SNPs and short indels. It was built using NGS reads from 3024 rice genomes, mapped into five reference assemblies. It consists of more than 42 million SNPs and more than 3 million short indels. We have developed a user-friendly web interface to access this database and will discuss several use cases how SNP-Seek helps rice researchers.

W811: Rice Functional Genomics

Florigen-Induced Transposon Silencing in the Shoot Apex during Floral Induction in Rice

Hiroyuki Tsuji, Kihara Institute for Biological Research, Yokohama, Japan

Floral induction is a crucial developmental step in higher plants. Florigen, a mobile floral activator that is synthesized in the leaf and transported to the shoot apex, was recently identified as a protein encoded by FLOWERING LOCUS T (FT) and its orthologs; the rice florigen is Heading date 3a (Hd3a) protein. The 14-3-3 proteins mediate the interaction of Hd3a with the transcription factor OsFD1 to form a ternary structure called the florigen activation complex on the promoter of OsMADS15, a rice APETALA1 ortholog. However, crucial information, including the spatiotemporal overlap among FT-like proteins and the components of florigen activation complex and downstream genes, remains unclear. Here, we confirm that Hd3a coexists, in the same regions of the rice shoot apex, with the other components of the florigen activation complex and its transcriptional targets. Unexpectedly, however, RNA-sequencing analysis of shoot apex from wild-type and RNA-interference plants depleted of florigen activity revealed that 4,379 transposable elements (TEs; 58% of all classifiable rice TEs) were expressed collectively in the vegetative and reproductive shoot apex. Furthermore, in the reproductive shoot apex, 214 TEs were silenced by florigen. Our results suggest a link between floral induction and regulation of TEs.

W812: Root Genomics

Dissecting the QTLome Governing Root System Architecture Features in Durum Wheat

Roberto Tuberosa, DipSA - University of Bologna, Bologna, Italy

Linkage (biparental) and association mapping were used for the dissection of the QTLome governing root system architecture (RSA) in seedlings of two recombinant inbred line populations and one association mapping panel of 183 elite durum wheat (*Triticum turgidum* L. var. *durum* Desf.), respectively. In total, 20 clusters of QTLs for root length and/or number as well as 30 QTLs for root growth angle (RGA) were evidenced. QTLs were mapped on a high-density tetraploid consensus map based on a transcript-associated Illumina 90K SNP assay developed for bread and durum wheat, thus allowing for an accurate cross-referencing of RSA QTLs between tetraploid and hexaploid wheat. Among the main QTL clusters for root length and number highlighted in this study, 15 overlapped with QTLs for multiple RSA traits reported also in bread wheat while out of 30 QTLs for RGA, only six colocated with QTLs previously reported in wheat. Based on their relative additive effects, allelic distribution in the AM panel and co-location with QTLs for yield and kernel weight, the RSA QTLs have been prioritized in terms of breeding value. Three major QTL clusters for root length and number and five QTLs for RGA appear particularly suitable for a possible deployment in marker-assisted selection and positional cloning.

W813: Root Genomics

Single Cell-Type Analysis Reveals Unique Patterns of DNA Methylation in the Root Meristem

Manuel Valdes, Duke University, Durham, NC

DNA methylation is an epigenetic marker of cell types, but the dynamics of DNA methylation changes between closely related cell types and its biological significance is unclear. Here, we report the genome-wide DNA methylomes, transcriptomes, and small RNA transcriptomes of six root cell types. We identified widespread cell type specific DNA methylomes, especially in CHH context. Columella genome is hypermethylated, accompanied by upregulation of RNA-directed DNA methylation related factors and 24 nt small RNA accumulation. Contradictory, nucleosome remodeler DDM1 required for DNA methylation is absent in columella. Low abundance of heterochromatin related factors indicates loss of heterochromatin in columella genome. We hypothesize that decondensed chromatin may allow access of RdDM factors to whole genome, producing excess 24 nt small RNAs in columella.

W814: Root Genomics

Exploring Root Morphological and Anatomical Plasticity Among Cereals to Enhance Adaptation to Water Limited Conditions

Krishna Jagadish, Kansas State University-Department of Agronomy, Manhattan, KS, Niteen Kadam, Wageningen University, Wageningen, Netherlands, Raju Bheemanahalli, International Rice Research Institute, Metro Manila, Philippines and Vara Prasad, Kansas State University, Manhattan, KS

Exploring genetic diversity in rooting morphology and anatomy provides tremendous opportunities for enhancing resilience to water-deficit stress conditions among field crops. Across the spectrum, we have cereals that are predominantly grown under fully flooded conditions (example rice) and some that are productive where water availability is significantly lower (example Sorghum). Our initial investigation included physiological, morphological and anatomical comparison of root-shoot dynamics, involving contrasting rice and wheat accessions, exposed to saturated and water-deficit stress conditions (60% Water Holding Capacity). Thicker roots and leaves and moderate tillering facilitated wheat to conserve water more effectively under water-deficit condition, compared to rice. Additionally, plasticity in stele and xylem diameter and xylem number along the root length in wheat cultivars facilitated efficient use of available moisture under water-deficit stress condition. A similar phenomena in root anatomical plasticity was not seen in popular rice cultivar (IR64) adapted to flooded conditions but some indications of plasticity were recorded in upland rice cultivar (N22). Based on the above findings two different studies were initiated (i) Comparative analysis of 23 different wild rice accessions with popular rice cultivars and other upland cereals including wheat, sorghum and maize (ii) Genome Wide Association Mapping for root morphological and anatomical parameters using ~290 diverse indica rice accessions exposed to saturated and water-deficit stress during maximum tillering stage. Opportunities for genetic modification of rice root morphology and anatomy, targeted root traits and chromosomal regions that can potentially induce greater resilience to water-deficit stress in rice cultivars will be discussed.

W815: Root Genomics

Has Selection for Heading Date Inadvertently Reduced Root Variation in European Wheat?

Kai Voss-Fels[1], Matthias Frisch[2], Lunwen Qian[1], Stefan Kontowski[3], Wolfgang Friedt[1], Sven Gottwald[1] and Rod Snowdon[1], (1)Department of Plant Breeding, Justus Liebig University, Giessen, Germany, (2)Justus Liebig University, Giessen, Germany, (3)W. von Borries-Eckendorf GmbH & Co. KG, Leopoldshöhe, Germany

Improving the roots of modern bread wheat varieties is considered to be one key factor in breaking yield barriers and improving adaptation to extreme abiotic stress conditions in the face of climate change. However, wheat roots have been insufficiently characterized to date, and further molecular investigations that develop diagnostic markers and thereby enable breeders a genomics-based selection of improved root systems are strongly required.

We performed comprehensive greenhouse phenotyping for basic seedling growth traits in an extremely diverse collection of 215 international wheat lines. Subsequent GWAS and haplotype network analysis identified rare haplotype variants, significantly associated with root mass, containing loci with high similarity to rice candidate genes for root traits. Strong linkage disequilibrium with molecular markers associated to heading date indicates an inadvertent co-selection of specific root variants during wheat breeding.

Our results may provide breeders the possibility for genomics-based exchange of co-localized haplotypes in order to maintain specific heading characters while maximizing root variation.

W816: Root Genomics

Individual vs. Combinatorial Effect of Elevated CO2 Condition and Drought Stress on Physiological and Molecular Response of *Glycine max* Roots

Marcio Alves Ferreira, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Climate changes increasingly threaten plant growth and productivity. Soybean (*Glycine max*) is one of the most important crops in the world and although responses of soybean to increase in atmospheric [CO_2] have been previously studied, there is no report of root molecular responses of this plant under elevated [CO_2] (E[CO_2]) nor the combination of E[CO_2] and drought. In this study, we evaluated the individual and combinatory effect of E[CO_2] and drought on physiology and root molecular responses in soybean. Plants were grown hydroponically in open-top chambers under ambient and E[CO_2] and drought stress was applied followed by RNA-seq analyses. Plants growing under E[CO_2] had photosynthesis increased which result in higher biomass, plant height and leaf area. E[CO_2] decreases transcripts levels of genes related to iron uptake and transport, antioxidant activity, secondary metabolism and defense and stress responses in roots. Both physiological and transcriptomic analysis data showed that E[CO_2] may mitigate in some extend the negative effects of drought. However, the identification of genes that are modulated by the interaction effect CO_2:drought may draw attention to unexpected consequences to soybean yields. Our data provide a distinctive overview of the molecular responses of soybean roots to the global climate changes.

W817: Root Genomics

Iron Response Genes in Rice, Genome Landscape and Function

Antonio Costa De Oliveira, Universidade Federal de Pelotas, Pelotas-RS, RS, Brazil

Rice is an important crop in Brazil, which is the largest producer outside Asia. The majority of rice produced in Brazil is under irrigation by flooding, which enables high yields due to its ability to better control weeds and to promote a more stable production environment. The water layer forms an anaerobic environment in which bacteria reduce organic compounds and elements. Among these changes, iron is reduced to ferrous iron, which is readily available to rice plants and, depending on its excess availability in the soil, can become toxic. Our group has been studying iron homeostasis genes and the response of tolerant and sensitive genotypes to iron excess. Microarrays and later RNASeq experiments have revealed an interesting pattern of response in genes and retrotransposons to iron excess. The progress in this understanding of iron homeostasis in rice is discussed.

W818: Sequencing Complex Genomes

IWGSC Whole Genome Shotgun Sequencing of Chinese Spring to Complement the High Quality BAC-Based Sequencing of All 21 Wheat Chromosomes

Curtis J Pozniak[1], Andrew G. Sharpe[2], Jesse Poland[3], Assaf Distelfeld[4], Jan Dvorak[5], Ming-Cheng Luo[5], Dr. Gil Ronen[6], Mike Thompson[7], Kellye Eversole[8], Jane Rogers[9], The International Wheat Genome Sequencing Consortium[10] and **Nils Stein**[11], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada, (3)Kansas State University, Manhattan, KS, (4)Tel Aviv University, Tel-Aviv, Israel, (5)Department of Plant Sciences, University of California, Davis, Davis, CA, (6)NRGene, Ness Ziona, Israel, (7)Illumina, Inc, River Falls, WI, (8)Eversole Associates, Bethesda, MD, (9)International Wheat Genome Sequencing Consortium, Cambridge, United Kingdom, (10)IWGSC, Lee's Summit, MO, (11)Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany

Wheat is one of the world's most important food crops and provides >20% of the protein and calories for the world's population. Breeding and agronomic improvements have achieved a linear increase in global food production but 70% more food will be required by 2050, which represents a 38% increase over historical rates. To keep pace, improved wheat cultivars are needed at a time when yield gains are stagnating in many parts of the world due to increased pest and disease pressure and extreme weather patterns linked to climate change. Of all major crops, wheat is the only major crop that lacks an ordered genome sequence, which has slowed identification of genes underlying phenotypic expression of agriculturally important traits.

The International Wheat Genome Sequencing Consortium (IWGSC) was established in 2005 as an international, collaborative consortium with the goal of producing a high quality reference sequence of the 17 Gbp allohexaploid (2n = 6x = 42), bread wheat genome, Chinese Spring. The IWGSC has followed an approach to develop separate physical maps, draft sequences, and high quality sequencing of the Minimum Tiling Paths (MTP) of mapped BAC clones for each chromosome. Recent innovations in Illumina sequencing technology and NRGene's computational genomics have enabled the IWGSC to produce rapidly a whole genome shotgun sequence of the Chinese Spring genome to complement the ongoing BAC-based sequencing thus aiming to accelerate the completion of the gold standard reference sequence. The latest results produced by the IWGSC will be presented.

W819: Sequencing Complex Genomes

High Quality Draft Genomes of *Medicago truncatula* & *Gossypium herbaceum* using NGS, Dovetail & BioNano Technologies

Thiruvarangan Ramaraj[1], Nicolas Devitt[1], Diego A. Fajardo[1], Karen M. Moll[1], Aaron Sharp[2], Kevin AT Silverstein[3], Jason R. Miller[4], Josh Udall[2], Nevin D. Young[5] and Joann Mudge[1], (1)National Center for Genome Resources (NCGR), Santa Fe, NM, (2)Brigham Young University, Provo, UT, (3)Supercomputing Institute, University of Minnesota, Minneapolis, MN, (4)J. Craig Venter Institute, Rockville, MD, (5)Department of Plant Pathology, University of Minnesota, St. Paul, MN

Here we present a high quality draft genome of *Medicago truncatula* (HM340), a model plant for studying legume symbioses and *Gossypium herbaceum* (cv. Wagad), a *A*-genome diploid from the Old World with genome sizes of approximately 0.5 & 1.7 Gbp respectively using a multi-platform sequencing strategy (Illumina HiSeq, PacBio RS II, Dovetail Genomics, and BioNano Whole Genome Maps). The multi-platform sequencing strategy will help immensely in achieving high quality *de novo* reconstructions of genomes, which requires accurate long-range contiguity. Using hybrid sequencing and analysis will not only provide high quality reference genomes which is a necessary foundation for several biological discoveries but also will aid in assessing recent technologies such as Dovetail Genomics and BioNano Genomics and also serve as a model to the plant genomics community who has an interest in using multi-platform sequencing technologies for *de novo* genome sequencing.

W820: Sequencing Complex Genomes

Spotting the Difference: Comparing the Genome of Corymbia with its Larger Cousin *Eucalyptus grandis*

Mervyn Shepherd[1], Kerrie W. Barry[2], Abdul Baten[3], Jakob Butler[4], Jules S. Freeman[5], Agnelo Furtado[6], Dario Grattapaglia[7], Adam Healey[6], Robert J. Henry[8], **Graham J King**[3], David Lee[9], Brad M. Potts[5], Jeremy Schmutz[10], Orzenil B. da Silva Junior[11], Blake Simmons[12] and Rene Vaillancourt[4], (1)Southern Cross University, Lismore, NB, Australia, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)Southern Cross Plant Science, Southern Cross University, Lismore NSW, Australia, (4)University of Tasmania, Hobart, Australia, (5)School of Plant Science and NCFFI, University of Tasmania, Hobart, Australia, (6)University of Queensland, Brisbane, Australia, (7)Plant Genetics Lab - EMBRAPA Genetic Resources & Biotechnology, Brasília, Brazil, (8)University of Queensland/QAAFI, Brisbane, Australia, (9)University of Sunshine Coast, Sippy Downs, Australia, (10)Hudson Alpha, Huntsville, AL, (11)EMBRAPA Genetic Resources and Biotechnology, Brasília, Brazil, (12)Joint Bioenergy Instutute, San Francisco, CA

Corymbia is a sister genus to *Eucalyptus* with *C. citriodora* (spotted gum) having a prominent role in forestry. In the past few decades it has been valued for timber, essential oil and energy production in the drier subtropical regions of Australia, China, India and Brazil.

The *Corymbia* Genome Consortium is generating a high quality genome sequence anchored to a dense genetic map to underpin evolutionary studies, germplasm management and accelerated breeding by genomic technologies. We have generated initial assemblies and a dense genetic map to facilitate comparative genome studies, particularly with the larger reference *Eucalyptus grandis* genome. The resources we have developed will facilitate transfer of genomic resources between the taxa, and broaden our understanding of the evolutionary drivers of genome variation within the eucalypt group. The *Corymbia* genome (reported as 370 Mb) is ~40% smaller relative to *E. grandis*. We will report progress on our understanding of the basis of this difference in genome size within the eucalypts, where the base chromosome size of 11 is remarkably prevalent. More specifically, we will present statistics of *de novo* assembly efforts using Illumina short reads, PacBio long reads and hybrid approaches that combine data from the two platforms. We demonstrate that hybrid assembly strategies even with low coverage PacBio data have already enabled us to generate a draft assembly of considerably improved quality for a complex genome. The *Corymbia* draft genome assembly will be released as an open-source platform and we are currently seeking additional partners.

W821: Sequencing Complex Genomes

Drafting the Kensington Pride Mango Genome

Natalie Dillon[1], **David Innes**[2], Yao Ming[3], Xiaodong Fang[4], Xuan Li[3], Xiaolei Gao[3], Ru Lin Zhan[5], Wu HongXia[5], Prasad Bajaj[6], Ian Bally[1], Alok Kumar[7] and Rajeev K Varshney[6], (1)Queensland Department of Agriculture and Fisheries, Mareeba, Australia, (2)Queensland Department of Agriculture and Fisheries, Dutton Park, Australia, (3)Beijing Genome Institute, Shenzhen, China, (4)BGI-Shenzhen, Shenzhen, China, (5)Subtropical Crops Research Institute, Zhanjiang, China, (6)ICRISAT, Hyderabad, India, (7)Horticulture Innovation Australia, Melbourne, Australia

The mango industry in Australia is worth in excess of $150 million annually with the Kensington Pride (KP) cultivar capturing 60% of the domestic market. Valued by consumers for desirable taste and colour characteristics, KP has been used extensively as a parent in the Department of Agriculture and Fisheries' (Queensland, Australia) mango breeding program with over 400 hybrid trees sharing KP as the male parent. In order to gain a better understanding of Australia's most significant mango variety, Horticulture Innovation Australia had led an international collaboration between the Queensland Department of Agriculture and Fisheries (Australia), the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT, India) and the Beijing Genomics Institute (China) to sequence the KP genome. Preliminary *de novo* assembly of illumina short read sequence data suggests that the KP genome is highly heterozygous and has an estimated genome size of 407 Mb. As refinements and additional sequence data are added to the assembly, a more complete picture of the mango genome will be elucidated.

W822: Sequencing Complex Genomes

Multiplex Sequencing of Bacterial Artificial Chromosomes for Assembling Complex Plant Genomes

Martin Mascher, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany

Hierarchical shotgun sequencing remains the method of choice for assembling high-quality reference sequences of complex plant genomes. The efficient exploitation of current high-throughput technologies and powerful computational facilities for large-insert clone sequencing necessitates the sequencing and assembly of a large number of clones in parallel. We developed a multiplexed pipeline for shotgun sequencing and assembling individual bacterial artificial chromosomes (BACs) using the Illumina sequencing platform. We illustrate our approach by sequencing 668 barley BACs (*Hordeum vulgare* L.) in a single Illumina HiSeq 2000 lane. Using a newly designed parallelized computational pipeline, we obtained sequence assemblies of individual BACs that consist, on average, of eight sequence scaffolds and represent >98 % of the genomic inserts. Our BAC assemblies are clearly superior to a whole-genome shotgun assembly regarding contiguity, completeness and the representation of the gene space. Our methods may be employed to rapidly obtain high-quality assemblies of a large number of clones to assemble map-based reference sequences of plant and animal species with complex genomes by sequencing along a minimum tiling path.

W823: Sequencing Complex Genomes

Rice Genome Sequences Explain the Evolution and Domestication of *japonica* and *indica* Rice

Marta Brozynska[1], Agnelo Furtado[2] and **Robert J. Henry**[2], (1)Queensland Alliance for Agriculture and Food Innovation, Brisbane, QLD, Australia, (2)University of Queensland/QAAFI, Brisbane, Australia

The wild A genome species represent an effective gene pool for rice. Sequencing of A genome species has allowed analysis of the relationships with this group. A whole chloroplast genome based phylogeny has now been compared with one based upon nuclear genome sequences. Wild taxa from Northern Australia have been included in these analyses for the first time. Compared with domesticated rice, some South American and African taxa had the most divergent chloroplast genomes and some Australian taxa the most divergent nuclear genomes. The domesticated rice (*O. sativa* japonica) nuclear genome shows close relationship with that of *O. rufipogon*confirming this as the progenitor of domesticated japonica. A wild rice from northern Australia was the closest perennial to indica. This supports domestication of indica from a separate gene pool found in Northern Australia that may still be present in Asia.

W824: Sex Chromosomes and sex determination

Evolutionary Mechanisms Underling the Gene Traffic out of the X Chromosome

Manyuan Long, University, Chicago, IL

The gene traffics between the sex chromosomes and autosomes are a widely observed pattern of new gene evolution in mammals and insects. Through extensive discussion and efforts made about the possible mechanisms involved, progresses have been made in understanding the evolutionary forces operating on the process. However, most proposed models were able to explain only one or a few facets of the gene traffics in only one or a few species. Is the pattern of new gene evolution formed by different mechanisms or is it shaped by the same force of evolution? The reported gene traffics in various organisms will be analyzed and compared for seeking an evolutionary force in common that is operating on these organisms.

W825: Sex Chromosomes and sex determination

Evolution of Genetic Sex Determination in Fish

Yann Guiguen, INRA-SCRIBE, Rennes, France

Fish show a great variety of sex determination mechanisms, which in the case of genetic sex determination is linked to a similarly high variability of sex chromosome differentiation. But this diversity does not follow any obvious phylogenetic pattern. To obtain a better understanding of the biological meaning of the diversity of sex determination and the mechanisms driving sex chromosome evolution we are attempting to decipher the molecular basis of the primary sex determination mechanisms and the structure and genetic organization of sex chromosomes across a broad diversity of rayfin fish. On the one hand we are analyzing a broad collection of species that represent major branches of the fish tree of life and on the other hand we focus on closely related species within branches of the phylogenetic tree (Salmoniformes, Esociformes, Danios, Poeciliids). We use high throughput marker mapping in 40 species as well as transcriptomics and genome sequencing to delineate sex-specific chromosomal regions and to identify candidate sex determining genes. These strategies already led to the identification of sex-specific markers, allowing delineating the extent of recombination suppression, and candidate sex determining genes in some species. Examples taken from Salmoniformes and its sister taxonomical clade i.e, the Esociformes will be described.

W826: Sex Chromosomes and sex determination

Identifying the Sex Determination Genes in Asparagus

Alex Harkess, Department of Plant Biology, University of Georgia, Athens, GA and Jim Leebens-Mack, University of Georgia, Athens, GA

Garden asparagus (Asparagus officinalis) is a dioecious species with a recently evolved, homomorphic sex chromosome pair which makes it ideal for studying the earliest events in sex chromosome evolution. A proposed evolutionary path from hermaphroditism to dioecy and a sex chromosome pair would involve the origin of a Y chromsome through cessation of recombination between a suppressor of female function and a promoter of male function. We have explored this hypothesis in garden asparagus by genetically mapping sex determination to a small (<2Mb) non-recombining region on the proto-Y chromosome that actively differentiates males (XY) from females (XX). We have identified two independent male-to-hermaphrodite mutants that implicate a single gene in this non-recombining region on the Y as responsible for dominantly interrupting pistil development. Anther development is not affected in these mutants. To identify the gene(s) responsible for anther sterility in XX females, we assessed gene expression profiles in male and female spears and identified known anther development genes that exhibit male-biased expression. A small subset of these genes do indeed map to sex linked scaffolds in the asparagus genome assembly, but their role in gender determination is still under investigation. Nonetheless, by identifying a female suppressor gene that does not influence anther development, these results support the hypothesis that at least two genes are necessary in the conversion of an autosome to a sex chromosome pair.

W827: Sex Chromosomes and sex determination

Evolution of a Flexible Sex Determination System in Polyploid Persimmon

Takashi Akagi[1], Isabelle M. Henry[2], Ryutaro Tao[1] and Luca Comai[3], (1)Graduate School of Agriculture, Kyoto University, Kyoto, Japan, (2)University of California, Davis, CA, (3)Plant Biology and Genome Center, UC Davis, Davis, CA

Epigenetic regulation adds a flexible layer to genetic variation, potentially enabling long term, but reversible cis-regulatory changes of an allele while maintaining DNA sequence. Here we present evidence for the establishment of an epigenetic switch regulating sex determination in persimmon. Specifically, we are documenting the role of DNA methylation in sex-determination in Oriental persimmon (*Diospyros kaki*). This regulation leads to the formation of both male and female flowers (monoeocy), in genetically male individuals. In these Y-chromosome carrying hexaploid persimmon trees, the gene encoding the female-determining factor *MeGI* exhibits 5' methylation and repression in buds that form male flowers but not in developing female flowers or any other tissue. Treatment with a demethylating agent could result in the reversion of male developing buds to form semi-feminized flowers, while it did not affect female flowers. This epigenetic control of flower development in hexaploid persimmon differs from the genetic determination observed in the closely related diploid persimmon. There, male flower development is based on the repression of *MeGI* by the Y-encoded male-determining factor *OGI*. In hexaploid persimmon, *OGI* expression is undetectable, possibly due to the insertion of a transposon in the 5' region of its promoter. This transposon insertion is conserved in all Y-carrying *D. kaki* cultivars, suggesting either natural selection or a strong bottleneck for the silenced *OGI* gene during the establishment of this species. Adaptive scenarios involving the relationship between this epigenetic plasticity and polyploid evolution will be discussed.

W828: Sex Chromosomes and sex determination

Whole Genome Sequencing and High Density Genetic Maps in Pistachio Reveal a Large Non-Recombining Region of Sex Chromosomes

Salih Kafkas[1], Haibao Tang[2], Rafael Navajas-Pérez[3], Hakan Ozkan[1], Andrzej Kilian[4], Ray Ming[5], Mortaza Khodaijman[1], Elmira Ziya Motalebipour[1], Murat Guney[1], Hayat Topcu[1], Ebru Kafkas[1], Nergiz Çoban[6], Hatice Gozel[6], Francisca Robles[3], Roberto de la Herrán[3], Carmelo Ruiz Rejón[3], Jason Carling[4], Jie Song[4] and William Wadlington[5], (1)University of Cukurova, Adana, Turkey, (2)University of Arizona, Tucson, AZ, (3)Universidad de Granada, Granada, Spain, (4)Diversity Arrays Technology Pty Ltd, Canberra, Australia, (5)University of Illinois at Urbana-Champaign, Urbana, IL, (6)Pistachio Research Institute, Gaziantep, Turkey

The genus *Pistacia* belongs to Anacardiaceae family, and has about 11 species. *Pistacia vera* is the only cultivated species in the genus. The 'Siirt' cultivar is one of the most commercially important cultivars in Turkey, and has also a highly heterozygous genome that presents challenges during the genome assembly. We obtained ~100x PacBio data and used along with 305x Illumina data for genome assembly using a hybrid, multi-stage assembly process. The final assembly contained 1,787 scaffolds with a total length of 614 Mbp, N50 contig size of 680Kb, and N50 scaffold size of 1.51 Mb. The unigene coverage was estimated to be 96.2%. The pistachio genome have been annotated to contain 61K genes, which included TE-related genes. The scaffolds were explored for repetitive elements, and 55% of them were assigned to 309 repetitive clusters. In addition to whole genome sequencing, we used three F1 segregating populations for genetic mapping by SNP markers and obtained a consensus map comprised of 11.2 K markers. This map as well as individual maps were used to anchor individual scaffolds onto chromosomes, and 85% of the scaffolds were located into the 15 chromosomes of pistachio. We also found that the largest linkage group LG1 (chr1; about 54 Mb) is the ZW sex chromosomes in pistachio, and has about 30 Mb region with no recombination, more than 50% of the W chromosome. Whole genome sequence of pistachio, along with well-characterized sex chromosomes, provide pistachio breeders new tools and methodologies to accelerate cultivar and rootstock breeding programs.

W829: Sex Chromosomes and sex determination

A Cucurbit Androecy Gene Reveals How Unisexual Flowers Develop and Dioecy Emerges

Adnane Boualem, INRA-IPS2, Orsay, France

Understanding the evolution of sex determination in plants requires identifying the mechanisms underlying the transition from monoecious plants, where male and female flowers coexist, to unisexual individuals found in dioecious species. We show that in melon and cucumber, the *androecy* gene controls female flower development and encodes a limiting enzyme of ethylene biosynthesis, ACS11. *ACS11* is expressed in phloem cells connected to flowers programmed to become female, and *ACS11* loss-of-function mutants lead to male plants (androecy). *CmACS11* represses the expression of the male promoting gene *CmWIP1* to control the development and the coexistence of male and female

flowers in monoecious species. Because monoecy can lead to dioecy, we show how a combination of alleles of *CmACS11* and *CmWIP1* can create artificial dioecy.

W830: SGN and RTB Databases: Genomics and Breeder Tools.

Introduction to SGN and Genomics Tools

Lukas Mueller, Boyce Thompson Institute for Plant Research, Ithaca, NY

The Sol genomics network (SGN) hosts genomics and phenotype data of the solanaceaes and related species. In this presentation we will cover the tools to access and visualize the data in the database.

W831: SGN and RTB Databases: Genomics and Breeder Tools.

solGS: A Web-based Solution for Genomic Selection

Isaak Y. Tecle, Boyce Thompson Institute for Plant Research, Ithaca, NY

Genomic selection, due to its reliance on dense genome-wide markers and statistical complexity, presents significant challenges in data management, analysis and sharing results. solGS, a web-based tool, meets these challenges; it has a database to store phenotype and genotype data and an intuitive web-interface for statistical analyses. It uses RR-BLUP for the statistical modeling and GBLUP method for breeding values estimation. It performs also descriptive statistics, population structure, phenotypic and genetic correlations, and selection index analyses. It visualizes data in interactive plots on the browser. solGS is, currently, used by the NextGen Cassava Breeding Project (http://nextgencassava.org) and implemented on http://cassavabase.org/solgs. GS breeders can adapt the tool for any organism.

W832: SGN and RTB Databases: Genomics and Breeder Tools.

The Breeder Toolbox, a Versatile Kit to Manage Phenotypes and Breeding Trials in Multiple Crops

Guillaume J. Bauchet[1], Isaak Y. Tecle[1], Naama Menda[1], Alex C. Ogbonna[1,2], Bryan Ellerbrock[1], Nicolas Morales[1] and Lukas Mueller[1], (1)Boyce Thompson Institute for Plant Research, Ithaca, NY, (2)National Root Crops Research Institute (NRCRI), Umuahia, NY, Nigeria

The Solgenomics (SGN) platform has developed a broad range of solutions and tools to help the scientific community to manage the ever increasing amount of genomic and phenomic data.
In this perspective, the Breeder Toolbox was designed to answer community's needs in the daily management of plant breeding programs. This support includes both field and computer work aspects.
Solutions for germplasm (accessions and trait search, pedigree construction), trial design (field maps, crossings), phenotypic and genotyping data collection (fieldbook, barcodes) and genotyping are now available in multiple crops including Solanaceae (tomato, potato, pepper, eggplant) but also root and tuber crops such as cassava, sweet potato or yam.
To address different community needs, dedicated web portals (solgenomics.net, cassavabase.org, sweetpotatobase.org, yambase.org) include common tools (i.e. accession search) and specific features answering each crop breeding requirement (i.e. breeding cycles).
Altogether, the Breeder Toolbox is a versatile tool that efficiently share and retrieve data with and from other SGN analytical components such as the genomic selection module, SolGS.

W833: Small RNA

Ten Years of Advances in Small RNA Biology

Xuemei Chen, University of California, Riverside, CA

W834: Small RNA

Insect Control by the Expression of Long dsRNA in Plastids

Jiang Zhang, Hubei University, Wuhan, China

Double-stranded RNAs (dsRNAs) targeted against essential genes can trigger a lethal RNAi response in insect pests. However, although expression of dsRNAs targeted against insect genes in transgenic plants has impaired growth and development, complete protection of the plants and efficient killing of the insects were not achieved due to the presence of an endogenous RNAi pathway in plants that processes dsRNAs into short interfering RNAs (siRNAs). We found that long dsRNAs can be stably produced in plastids, a cellular compartment that appears to lack an RNAi machinery. When expressed from the chloroplast genome, dsRNAs accumulated to as much as 0.4% of the total cellular RNA. Transplastomic potato plants producing dsRNAs targeted against the β-actin gene of the Colorado potato beetle, a notorious agricultural pest, were efficiently protected from herbivory and were lethal to its larvae. Thus, plastid-expressed long dsRNAs can provide full crop protection without chemical pesticides and without synthesis of foreign proteins in the plant. Shifting the target of transgesis from the nucleus to the plastid removes the major hurdle on the way to exploiting transgenically delivered RNAi for efficient crop protection in the field.

W835: Small RNA

Small RNA-Producing Genes in Plants: Improved Methods and Novel Discoveries

Michael Axtell, Penn State University, University Park, PA

Plants produce an extraordinary variety of regulatory small RNAs, which include microRNAs (miRNAs) and short interfering RNAs (siRNAs). miRNAs and siRNAs are a major factor in plant gene regulation and affect nearly every phenotype of interest. A given plant genome produces miRNAs from many hundreds of genes, and siRNAs from many thousands of genes. However, the methods for finding and annotating these genes have been uneven, and the current state of miRNA and siRNA gene annotation is poor. I will describe our novel, open-source methods for miRNA and siRNA gene discovery using high-throughput small RNA-seq data. I will then describe how we have used these methods to make novel small RNA discoveries in several plant species, including *Arabidopsis*, maize, *Physcomitrella patens*, and most recently in several parasitic plant species.

W836: Small RNA

A Genome-wide Atlas of MicroRNA Primary Transcript Structure

Tsung-Cheng Chang, University of Texas Southwestern Medical Center, Dallas, TX

MicroRNA (miRNA) expression is dynamically regulated during development, across tissues, and in various human diseases. Nevertheless, a major bottleneck in the elucidation of mechanisms that control miRNA abundance is the currently incomplete annotation of primary miRNA (pri-miRNA) gene structures. While a subset of miRNAs are hosted in protein-coding genes, the majority of pri-miRNAs are transcribed as poorly-characterized noncoding transcripts. Due to the efficiency of DROSHA processing, the abundance of pri-miRNAs is very low at steady-state. Therefore, elucidation of pri-miRNA structure has remained a significant challenge. To address this problem, we developed an experimental and computational approach that allows rapid transcriptome-wide mapping of pri-miRNA structures. By performing deep RNA-seq in cells expressing a dominant-negative DROSHA mutant protein, we demonstrated dramatic enrichment of intact pri-miRNAs, resulting in much greater coverage of these transcripts compared to standard RNA-seq. We also evaluated multiple assembly algorithms, ultimately demonstrating that StringTie, a software package that we recently developed, outperforms other existing assembly tools for this application. We applied our novel pri-miRNA annotation strategy to a panel of human and mouse cell lines of diverse origins, thereby significantly improving the existing annotation of conserved mammalian miRNA genes. Based on these new assemblies, we uncovered unanticipated features and new potential regulatory mechanisms, including unexpected links between pri-miRNAs and distant protein coding genes, alternative pri-miRNA splicing, and alternative promoter usage that can produce transcripts carrying subsets of miRNAs encoded by polycistronic clusters. These results provide a valuable resource for the study of mammalian miRNA regulation.

W837: Small RNA

Establishment of heritable transposon-like transgene silencing in *Arabidopsis* is disrupted in large multigenic T-DNAs with GC-rich coding sequences

Tzuu-fen Lee, University of Delaware, Newark, DE

Variability and eventual loss of transgene expression due to silencing is a significant challenge for plant genetic engineering, especially for large, complex transgene insertions. We have characterized in detail the silencing mechanisms that occur when a suite of transgenes encoding a microalgal polyunsaturated fatty acid (PUFA) synthase system is introduced into *Arabidopsis*. Certain transgene constructs using coding sequences (CDS) with a plant codon bias underwent aggressive silencing, similar to that of active transposable elements (TEs). Silencing was mitigated by diversifying regulatory sequences and was prevented when GC-rich native microalgal CDS were substituted for canola-biased transgenes. This was accompanied by reduced accumulation of 21/22-nt siRNAs, low levels of 24-nt siRNAs and sparse (gene-body-like) CDS DNA methylation, indicating that the transition from modest transcript degradation to heritable silencing was suppressed by using high GC transgenes. These observations provide a strategy for countering transgene silencing in genetic engineering of complex plant traits.

W838: Small RNA

miR160 Action and Transcriptome Profiles Suggest that Root Nodules might have Evolved from Shoot Lateral Organs

Senthil Subramanian, South Dakota State University, Brookings, SD

We used global gene and microRNA (miRNA) expression profiles to identify potential candidates that might play key roles in soybean nodule development. Comparative expression analysis of miRNAs and their potential target genes identified key miRNA-target pairs potentially involved in nodule development. Of particular interest to us was miR160 that regulates the levels of repressor auxin response transcription factors belonging to the ARF10/16/17 family. We observed an inverse expression pattern between miR160 and the majority of its validated targets. We evaluated the functional significance of miR160 levels through ectopic expression and suppression assays combined with hormone response assays, fluorescent sensor microscopy, and hormone rescue assays. Results from these experiments indicated that miR160 dictates developmental stage-specific balance between auxin and cytokinin to direct proper nodule formation and maturation in soybean. Interestingly, both miR160 activity and auxin-cytokinin balance required during nodule formation is similar to that of shoot axillary meristems and opposite to that of lateral root meristems. In addition, members of genes families associated with shoot axillary meristem formation were enriched in nodules, but not lateral roots in our transcriptome datasets. These observations provide strong support to the hypothesis that nodule formation might have evolved by recruiting developmental and hormonal signaling pathways from shoot axillary bud formation.

W839: Solanaceae

***S. verrucosum*, a Wild Mexican Potato As a Model Species for a Plant Genome Assembly Project**

Pirita M Paajanen[1], Michael Giolai[1], Walter Verweij[2], Dave Baker[1], Bernardo Clavijo[2], Gonzalo Garcia[2], Elena Lopez Girona[3], Katie Baker[3], Ingo Hein[3], Glenn Bryan[3] and Matt Clark[2], (1)The Genome Analysis Centre, Norwich, United Kingdom, (2)The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland, (3)The James Hutton Institute, Dundee, United Kingdom

In this project we sequence an inbred self-compatible tuber-bearing wild Mexican diploid potato species *S. verrucosum*. Using this genome we benchmark the latest Illumina, PacBio long reads, BioNano Genomics optical maps and Dovetail artificial Hi-C data along with Bioinformatics tools to generate scaffold N50s in the multi-megabase range.

We extracted DNA from fresh leaf tissue of a selfed line, from which various Illumina libraries constructed. Analysis of the k-mer profiles indicates that this *S. verrucosum* accession is highly homozygous as expected. Illumina large insert PCR-free, and DISCOVAR libraries were used for paired end data, Illumina Nextera mate pair libraries for scaffolding, with PacBio long reads for gap filling. The Illumina assemblies were further scaffolded using a Dovetail Genomics cHiCago™ in vitro Hi-C library followed by HiRise™ scaffolding. We are currently using a PacBio only *de novo* assembly, and additionally BioNano Genomics optical map based hybrid assemblies.

Benchmarking the assemblies consists of assessing the presence of CEGMA core eukaryotic genes, local assembly accuracy will be compared to BAC inserts, long range scaffolding accuracy based on synteny to *S. tuberosum* and *S. lycopersicum*, and correct assembly of repetitive regions as exemplified by R-gene clusters.

Generation of a genome sequence for *S. verrucosum* will open up further opportunities for its use as a genetic model, and for studying comparative genome structure within the Solanaceae.

W840: Solanaceae

Multiple *de novo* Genome Sequences of Pepper Provide Insights into Species Diversification in *Capsicum* spp.

Seungill Kim, Myung-Shin Kim and Doil Choi, Department of Plant Science, Plant Genomics and Breeding Institute and Interdisciplinary Program in Agricultural Genomics, Seoul National University, Seoul, South Korea

A diversity of *Capsicum* species has been derived from extensive genomic and phenotypic variations. Here, we report high-quality *de novo* genome sequences of two *Capsicum* species (*C. chinense* and *C. baccatum*) as multiple reference genomes of hot pepper. A total of 95 % of *C. chinense* (3.07 Gb of 3.21 Gb) and 76 % (3.22 Gb of 4.2 Gb) of *C. baccatum* genomes were assembled. Of these genome assemblies, 87.3 % (2.63 Gb) and 87.2 % (2.79 Gb) were anchored to 12 chromosome pseudomolecules, respectively. Comparative analysis with the existing pepper reference genome (*C. annuum*) revealed that the amount of ribosomal DNA sequences in *C. baccatum* was more than 10 fold compared to the other pepper genomes. Athila, a subgroup of *Gypsy* family, was excessively accumulated in *C. baccatum* and increased the genome size in this species compared to the others. Phylogenetic analysis revealed that speciation of *C. baccatum* and *C. chinense* from *C. annuum* had occurred at 1.7 and 1.1 million years ago, respectively. We identified correlations among proliferation of LTR retrotransposons, acceleration of gene expression change, gene duplication, and large genomic variations during the speciation of the pepper genomes. Our results support that rapid accumulation of LTR-retrotransposons during the speciation has contributed diversification of the three pepper genomes. The multiple pepper genomes will serve as important resources for comparative and population genomics as well as evolutionary studies of the genus *Capsicum*.

W841: Solanaceae

The Tomato Expression Atlas: A New Platform for Biological Discovery with Cell-Type Resolution

Jocelyn KC Rose, Cornell University, Ithaca, NY

Most biochemical and molecular studies involving the extraction of transcripts or proteins from plant organs use a homogenized amalgam of tissues and cell types. This approach limits insights into cell specialization, and lower abundance molecules that are present only in certain cell types are often diluted below the level of detection. There is therefore a critical ‘information void’ when it comes to annotating and presenting gene expression data. We have been addressing this challenge in the context of understanding the entirety of gene expression during tomato fruit development, by coupling RNA-seq analysis with laser capture microdissection (LCM), which allows the precise isolation of individual fruit cells/tissue types. In addition to resolving gene expression down to the level of cell/ tissue type, this approach has enabled: *(i)* the identification of previously unannotated genes, demonstrating the value of LCM as a tool for gene discovery; *(ii)* inferences regarding gene functions, based on the patterns of tissue- or cell type-related expression. We have also been developing computed tomography as a non-invasive imaging tool to create a 3D ‘virtual tomato’, which includes internal structures, to provide digital a scaffold upon which to present transcriptome, or other ‘omics’ data sets as a 4D display. All data will be publicly accessible in a new database, the Tomato Expression Atlas. This database includes a novel user interface with a correlation matrix that reveals patterns of co-expressed genes at an unprecedented level of spatiotemporal resolution, thereby optimizing the identification of functionally related suites of genes.

W842: Solanaceae

A Cascade of Arabinosyltransferases Controls Shoot Meristem Size in Tomato

Cao Xu, Cold spring harbor lab, cold spring harbor, NY

Shoot meristems of plants are composed of stem cells that are continuously replenished through a classical feedback circuit involving the homeobox *WUSCHEL (WUS)* gene and the *CLAVATA (CLV)*gene signaling pathway. In CLV signaling, the CLV1 receptor complex is bound by CLV3, a secreted peptide modified with sugars. However, the pathway responsible for modifying CLV3 and its relevance for CLV signaling are unknown. By studying mutants with enlarged meristems and branched inflorescences with extra flowers and floral organs, we show that tomato genes encoding arabinosyltransferases are essential for the CLV-WUS circuit. The most extreme mutant *fasciated inflorescence (fin)* is disrupted in a hydroxyproline O-arabinosyltransferase and can be rescued with arabinosylated CLV3. Chemically induced and CRISPR/Cas9 engineered mutations in the *FASCIATED AND BRANCHED 2 (FAB2)* and *REDUCED RESIDUAL ARABINOSE 3 (RRA3)*genes, encoding arabinosyltransferases predicted to extend arabinose chains, result in plants with fasciated and branched phenotypes, resembling but weaker than *fin*, indicating that CLV3 must be fully arabinosylated to maintain meristem size. Finally, we show that a mutation in CLV3 increased fruit size during domestication. Our findings uncover a new layer of complexity in the control of plant stem cell proliferation.

W843: Solanaceae

A Disease Resistance Locus on Potato and Tomato Chromosome 4 Evolves at Different Rates in Different Lineages

Dan Milbourne[1], Marialaura Destefanis[2], Istvan Nagy[3], Brian Rigney[1], Glenn Bryan[4], Karen McLean[4], Ingo Hein[4] and Denis Griffin[1], (1)Crops, Environment & Land Use Programme, Teagasc, Carlow, Ireland, (2)Pesticides, Plant Health & Seed Testing Laboratories, Department of Agriculture, Food and the Marine, Co. Kildare, Ireland, (3)Dept. of Molecular Biology and Genetics, Aarhus University, Slagelse, Denmark, (4)The James Hutton Institute, Dundee, United Kingdom

In plant genomes, NB-LRR based resistance (R) genes tend to occur in clusters of variable size in a relatively small number of genomic regions. R-gene sequences mostly differentiate by accumulating point mutations and gene conversion events. Potato and tomato chromosome 4 harbours a syntenic R-gene locus (known as the *R2* locus in potato) that has mainly been examined in central American/Mexican wild potato species on the basis of its contribution to resistance to late blight, caused by the oomycete pathogen *Phytophthora infestans*. Evidence to date indicates the occurrence of a fast evolutionary mode characterized by gene conversion events at the locus in these genotypes. A physical map of the *R2* locus was developed for three *Solanum tuberosum* genotypes and used to identify the tomato syntenic sequence. Functional annotation of the locus revealed the presence of numerous resistance gene homologs (*RGHs*) belonging to the *R2* gene family (*R2GHs*) organized into a total of 4 discrete physical clusters, three of which were conserved across *S. tuberosum* and tomato. Phylogenetic analysis showed clear orthology/paralogy relationships between *S. tuberosum R2GHs* but not in *R2GHs* cloned from *Solanum* wild species. This study confirmed that, in contrast to the

wild species *R2GH*s, which have evolved through extensive sequence exchanges between paralogs, gene conversion was not a major force for differentiation in *S. tuberosum R2GH*s, and orthology/paralogy relationships have been maintained via a slow accumulation of point mutations in these genotypes. *S. tuberosum* and *Solanum lycopersicum R2GH*s evolved mostly through duplication and deletion events, followed by gradual accumulation of mutations. Conversely, widespread gene conversion is the major evolutionary force that has shaped the locus in Mexican wild potato species. We conclude that different selective forces shaped the evolution of the *R2* locus in these lineages and that co-evolution with a pathogen steered selection on different evolutionary paths.

W844: Solanaceae

Genetic Diversity on Resistance to Tomato Yellow Leaf Curl Virus in Tomato

Zhe Yan[1], Myluska Caro Rios[1], Anne-marie, A Wolters[1], Samuel F. Hutton[2], Jay W. Scott[2], Ana Pérez-de-Castro[3], Maria J. Díez[3], Richard Visser[1], Junming Li[4] and **Yuling Bai**[1], (1)Wageningen UR Plant Breeding, Wageningen, Netherlands, (2)University of Florida, GCREC, Wimauma, FL, (3)Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV), Valencia, Spain, (4)Institute of Vegetable and Flowers, the Chinese Academy of Agricultural Sciences, Beijing, China

Tomato yellow leaf curl disease (TYLCD) is a major constraint in tomato (*Solanum lycopersicum*) production worldwide since 1980's. It is a viral disease caused by a cluster of virus species. Besides *Tomato yellow leaf curl virus* (TYLCV), the most invasive and best studied species, species of TYLCV-like viruses and strains have been identified and classified according to the description in the International Committee on Taxonomy of Viruses (ICTV). Controlling TYLCD is usually focused on reducing or avoiding the whitefly vector population by common practices of heavy spray of insecticides and/or the use of nets. Breeding TYLCV-resistant tomato cultivars is an economically and environmentally sustainable alternative. In the past, many wild tomato relatives have been screened for resistance to TYLCV. However, accession numbers were not always published and resistant sources were usually clarified only at species level, which led to redundant screenings for TYLCV resistance in different labs. This presentation will (1) summarize results of different screenings worldwide to illustrate genetic diversity on tomato resistance to TYLCV; (2) report on the progress and status of the identification and mapping of tomato genes (including *Ty-1* to *ty-6)* and their allelic variants; and (3) discuss the challenges of TYLCV introgression breeding (e.g. chromosomal rearrangements) and possibilities for durable and broad-spectrum geminivirus resistance.

W845: Somatic Genome

Somatic Genome Variation: Natural Attributes, Impacts on Human Health, and Implications for Agriculture

Xiu-Qing Li, Agriculture and Agri-Food Canada, Fredericton, NB, Canada

This introductory presentation will discuss the definition of a genome, a genome network, and a somatic genome, list the attributes of somatic genome variation in microorganisms, animals and plants, review the impacts of somatic genome variation on human health and plant development, and explore the use of somatic genome variation and related somatic breeding technologies for improving plants.

W846: Somatic Genome

Shattered but Alive – Surviving Mitotic Catastrophe during Genome Elimination

Ek Han Tan[1], Isabelle M. Henry[1], Ravi Maruthachalam[2], Keith Bradnam[1], Terezie Mandakova[3], Martin A. Lysak[3] and Luca Comai[1], (1)University of California, Davis, CA, (2)Indian Institute of Science Education and Research (IISER), Kerala, India, (3)CEITEC, Masaryk University, Brno, Czech Republic

Safeguarding the genome from instability is essential for maintaining the health and fitness of an organism. On the other hand, alterations to genomes provide an avenue for organisms to adapt to change. Small, gradual changes to DNA sequences have been well studied but more drastic alterations such as those derived from chromothripsis, the phenomenon in which a chromosome is shattered and reassembled incorrectly in a single step, are not well understood. We recently reported the occurrence of chromothripsis during genome elimination in *Arabidopsis thalian*a. We show that micronuclei are likely involved, providing a crucial mechanistic link between chromothripsis in plant and animal systems. In addition, we provide genetic evidence for the role of non-homologous end joining in repairing DNA breaks from shattered chromosomes. Using our system in *Arabidopsis*, the karyotypic novelty that results can also be passed on to the offspring. This enables the study of the molecular mechanisms of chromothripsis as well as its consequences across multiple generations in an organismal setting.

W847: Somatic Genome

Connecting DNA Repair Genes with Genome Divergence through Base Composition at Polymorphic Sites

Xianran Li[1], Michael Scanlon[2] and Jianming Yu[1], (1)Department of Agronomy, Iowa State University, Ames, IA, (2)Cornell University, Ithaca, NY

DNA base composition is a fundamental genome feature. However, the evolutionary pattern of base composition and its potential causes have not been well understood. Leveraging the wealth of genome sequence data, we characterized composition of nucleotides along the same DNA strand from different levels. Two consistent patterns were discovered. Base composition follows the individual-strand base equality rule at the genome, chromosome and polymorphic site levels. More intriguingly, clear separation of base-composition values calculated across polymorphic sites was consistently observed between basal and derived groups, suggesting common underlying mechanisms. With base-composition across polymorphic sites as a phenotype, genome scans with Human 1000 Genomes and HapMap3 data identified a set of significant genomic regions enriched with Gene Ontology terms for DNA repair. For three DNA repair genes (*BRIP1*, *PMS2P3* and *TTDN*), ENCODE data provided evidence for interaction between genomic regions containing these genes and regions containing the significant SNPs. Our findings provide insights into the mechanisms of genome evolution.

W848: Somatic Genome

The Increased Tolerance to Biotic and Abiotic Stresses and the Changes in Transcriptome and MicroRNA Activity after Chromosome Doubling in Paulownia Trees

Guoqiang Fan, Institute of Paulownia, Henan Agricultural University, zhengzhou, China
After doubling the chromosomes of various diploid *Paulownia fortunei* plants, the obtained tetraploid plants exhibited considerable differences in phenotypic traits and tolerance to both biotic and abiotic stresses from the original diploids. To understand the molecular mechanisms underlying these differences, we compared the small RNAome, degradome, transcriptome and proteomics between diploid and tetraploid *P. fortunei* by mRNA-seq, sRNA-seq degradome and isobaric tags for relative and absolute quantitation (iTRAQ) labeling techniques in the current study. Compared with the diploid plants (PF2), the tetraploid plants (PF4) showed up or down regulation in expression of 187 miRNAs and reverse patterns (PF4 vs PF2) of the expression activities of 25 miRNA-targeted genes. Among a total of 2355 proteins identified, 119 proteins were differentially expressed between diploid and tetraploid *P. fortune*. Finally, a total of 35 differentially expressed proteins related to plant growth and development were identified. Together, these results indicate that a large number of miRNAs can be differentially expressed between diploid and tetraploid *P. fortunei* plants and that both conserved and novel miRNAs can cleave their target mRNA sequences. This study, for the first time in *P. fortunei*, has characterized differentially expressed proteins and associated mRNAs after chromosome doubling and provided candidate genes for genetic improvement of Paulownia trees.

W849: Somatic Genome

Epigenome Reprograming during Fruit Development

Silin Zhong, The Chinese University of Hong Kong, Hong Kong, Hong Kong
DNA cytosine methylation is a conserved epigenetic mark present in many eukaryotes including plants, animals and fungi. It plays a vital role in maintaining genome integrity and controlling gene expression. Fleshy fruit ripening is an irreversible developmental process that the physiological and biochemical properties of the seed-bearing organ are altered to foster animal mediated seed dispersal. Tight regulatory oversight of this process is required to insure accurate and tissue-specific control of a developmental transition that would be highly detrimental if deployed in the wrong tissue or stage of fruit maturity. In the model fruit tomato, a whole-genome demethylation event during fruit ripening coincides with seed maturation have been identified. Demethylation occurs at the promoter of key fruit ripening genes such as those encoding transcription factors and enzymes associated with secondary metabolite production. This suggests that the relatively stable epigenome could be serving as a checkpoint to prevent ripening before seed maturation. However, whether this strategy is adopted by others remain largely unknown. We examined the epigenomes and transcriptomes of other climacteric fruits such as papaya, peach, banana, melon etc. We observed similar whole-genome demethylation events as the tomato one. To our surprise, the genes associated with differentially methylated regions (DMRs) in these fruits are drastically different from the differentially methylated tomato genes. For example, it is believed that MADS-box (RIN), NAC-domain (NOR) and SPB-box (CNR) transcription factors have conserved roles in fruit development, and they are associated with promoter DMRs during tomato fruit ripening. Intriguingly, many of their homologs lack DMR in the promoter region. These suggest that the epigenome reprograming is indeed present in other fruit-bearing plant species, but their transcriptional regulatory network could be very different to the tomato one. Hence, the epigenome reprograming might have to target different set of genes to exert its influence on fruit ripening.

W850: Somatic Genome

Capturing Short Tandem Repeat Variation to Measure Genome Instability

Keisha Carlson, University of Puget Sound, Tacoma, WA
Genome instability is a hallmark of many cancers and may be a general marker for increased disease susceptibility. To measure genome instability with high resolution, we propose assessing somatic variation in short tandem repeats (STRs). STRs are highly mutable genomic elements that, due to their repetitive nature, are prone to mutation from strand slippage during DNA replication and misalignment during homologous recombination. Accurately genotyping STRs in a high-throughput manner remains challenging for several reasons, one of which is the high level of “amplification” stutter that confounds identification of true, biological variation. Here we present our method MIPSTR to address the technical challenges of genotyping repetitive DNA. Using single molecule DNA barcodes, MIPSTR allows us to accurately genotype somatic STR variation in additional to germline STR variation across many individuals. Applying MIPSTR to *Arabidopsis thaliana*, we find increased somatic STR variation in *msh2* mutants deficient in mismatch repair as well as in HSP90-reduced plants. The increased somatic STR variation in these plants correlates with increased genome instability and increased expressivity of EMS-induced mutations.

W851: Somatic Genome

Going the Distance: Challenges in Long-Range Gene Regulation

Michael W. Dorrity, University of Washington, Seattle, WA
Plants respond to environmental stimuli by tightly controlled changes in gene expression, requiring a dynamic regulatory network. Despite this need, a key mode of dynamic regulation, long-range activation or repression, appears to be absent from chromatin interaction datasets (3C, Hi-C) in *Arabidopsis thaliana.* The apparent lack of distal physical interactions among chromosomes suggests that *Arabidopsis* may not use typical enhancers, though they are frequent among other higher eukaryotes. We describe the use of an alternative method to identify and functionally characterize enhancers, STARR-seq. This method relies on simultaneously measuring the activity of a library of gene constructs, each containing a fragment of genomic DNA inserted into a transcribed sequence. If a fragment does not possess an enhancer element, the construct is expressed at low level due only to the minimal promoter driving transcription. If a fragment is able to confer distal activation, the gene is expressed at higher level and the functional enhancer sequence, codified in the transcript, increases in abundance. We have optimized the method for transient expression in plants by testing a known viral enhancer in the experimental context, and demonstrate the scale to which the method can leverage next-generation sequencing to pinpoint functional enhancers genome-wide. Further, we present an orthogonal approach to characterizing the functional significance of distal interactions by forced chromosome contacts (f-C). We foresee both methods being useful for addressing biological questions of gene regulation, as well as for application to emerging challenges in crop design.

W852: Sorghum/Millet

Welcome and Introduction

Yinghua Huang, USDA ARS, Stillwater, OK

W853: Sorghum/Millet

Genomic Characterization of a Core Set of the USDA-NPGS Ethiopian Sorghum Germplasm Collection

Hugo E. Cuevas, ARS - Tropical Agriculture Research Station, Mayaguez, PR

The USDA-ARS National Plant Germplasm System (NPGS) preserves the largest sorghum germplasm collection in the world and includes 7,217 accessions from the center of diversity located in Ethiopia. This exotic germplasm has not been characterized on a genome-wide basis to better inform its conservation and utilization in research and breeding programs. Therefore, a representative core set of 374 Ethiopian accessions were phenotyped and characterized at the genomic level through genotyping-by-sequencing (GBS). A total of 148,476 single-nucleotide polymorphism (SNPs) markers distributed across the entire genome were identified, which over half of the SNPs were rare (frequency < 0.05). The genetic profile of each accession was unique (i.e. no duplicates), and 93% of the pairwise genetic distance among them ranged from 0.60 to 0.79. Population structure and cluster analysis separated the collection into three subpopulations with a high differentiation index (F_{ST} = 0.15). Phenotypic analysis confirmed these subpopulations based on agronomic and seed compositional traits. Cluster analysis with the sorghum association panel based on 60,962 SNPs determined that two of the Ethiopian subpopulations were not adequately represented in this panel. Indeed, the population structure of the NPGS Ethiopian collection can be exploited to screen this exotic germplasm. The large numbers of rare alleles indicates that this germplasm contains undiscovered and potentially useful alleles, but their discovery and characterization will require much additional effort. This genomic characterization provide a valuable resource for sorghum breeders and geneticists to effectively explore the potential of this highly diverse germplasm for sorghum improvement

W854: Sorghum/Millet

Genome-wide Annotation of Mutations in a Phenotyped Mutant Library Provides an Efficient Platform for Discovery of Causal Gene Mutations

Yinping Jiao[1], John Burke[2], Ratan Chopra[2], Gloria Burow[2], Junping Chen[2], Bo Wang[3], Chad Hayes[2], Yves Emendack[2], Doreen Ware[4] and Zhanguo Xin[5], (1)USDA-ARS/Cold Spring Harbor Laboratory, Lubbock, TX, (2)USDA-ARS, Lubbock, TX, (3)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (4)Cold Spring Harbor Laboratory/USDA-ARS, NY, NY, (5)USDA ARS, Lubbock, TX

Ethyl methanesulfonate (EMS) efficiently generates high-density mutations in genomes. Conventionally, these mutations are identified by techniques that can detect single-nucleotide mismatches in heteroduplexes of individual PCR amplicons. We applied whole-genome sequencing to 256-phenotyped mutant lines of sorghum (*Sorghum bicolor* L. Moench) to 16x coverage. Comparisons with the reference sequence revealed >1.8 million canonical EMS-induced G/C to A/T mutations, 22% of which were in genic regions, affecting >95% of genes in the sorghum genome. The vast majority (97.5%) of the induced mutations were distinct from natural variations. By applying the mutation database to phenotype analysis, we identified four causal gene mutations affecting drought tolerance, two mutations affecting heat tolerance, and two mutations affecting seed size that corresponded to previously reported seed size QTLs. Our results demonstrate that this collection of sequenced mutant lines can be used to efficiently discover new traits and their underlying causal mutations, thereby accelerating sorghum breeding.

W855: Sorghum/Millet

Biosynthesis, Regulation and Genetic Associations for Phytoalexins Induced during Sorghum-Colletotrichum interactions

Surinder Chopra, Pennsylvania State University, University Park, PA

W856: Sorghum/Millet

A Field-based High-throughput Phenotyping Platform to Discover the Genetic Architecture of Sorghum Biomass Yield Components Over Time

Maria G. Salas Fernandez, Yin Bao, Lie Tang and Patrick S. Schnable, Iowa State University, Ames, IA

High-throughput phenotyping technologies have emerged as a consequence of the need to obtain data at large scale, to increase accuracy and repeatability, and to phenotype plants over time for complex traits that could not be characterized by hand. The first succesful attempts to create high-throughput phenotyping systems were reported for laboratory or greenhouse settings or for field crops with small plant architecture. We have created a novel field-based self-propelled platform equiped with high resolution cameras that was specifically designed for a high biomass crop such as sorghum. This technology has been used to collect digital images during the entire season for 700 diverse sorghum lines in multiple environments. Data could be collected for a single location in a few hours and images were subsequently processed with appropriate algorithms to obtain plant architecture parameters such as plant height, stem diameter, panicle size and leaf angle. The accuracy of image-based algorithmically-derived data was demonstrated when compared with ground-truth measurements. These validated phenotypes will provide novel information about the dynamic changes in biomass production, growth rate and responses to variable environmental conditions during the growing season.

The phenotypic data generated in this project will be used to discover genes/SNPs associated with variation in plant architecture traits over time, could be further utilized for the genetic improvement of sorghum and represents a significant contribution to the emerging field of predictive phenomics.

W857: Sorghum/Millet

Advances in Genetics and Genomics of Foxtail Millet (*Setaria italica*) for Crop Improvement of Millets, Cereals and Bioenergy Grasses

Manoj Prasad, National Institute of Plant Genome Research, NEW DELHI, India

Foxtail millet (*Setaria italica*) is the oldest domesticated crop in the world (domesticated >8700 years ago) and it has been extensively grown in the semi-arid regions of Asia, Europe and Americas as food and fodder crop. Being a C_4 crop with genetic close-relatedness to several biofuel grasses, foxtail millet has been accentuated as a model crop. In view of its importance, U.S. Department of Energy Joint Genome Institute and

Beijing Genomics Institute have independently sequenced the genome of foxtail millet. The availability of draft genome sequence has advanced the genomics and genetics of this important crop resulting in development of large-scale genome-wide molecular markers and demonstration of their utility in genomics-assisted breeding, and delineation of several stress-responsive gene families for their molecular and biological roles in abiotic stress tolerance. In addition, several open access databases have been developed to cater these resources for crop improvement through structural and functional genomics. In view of this, the present talk will summarize the advances made in genetics and genomics of this $\mathbf{C_4}$ model crop and how this information could be translated for improvement of other millets, cereals and biofuel grasses.

W858: Sorghum/Millet

Application of Herbicide-Resistant Genes from Green Foxtail Millet in Foxtail Millet Breeding

Ruhong Cheng[1,2], Guoqing Liu[1,2], Zhigang Shi[1,2], Xueyan Xia[1,2] and Ting Zhang[1,2], (1)China National Millet Improvement Center, Shijiazhuang, China, (2)Institute of Millet Crops, Hebei Academy of Agricultural & Forestry Sciences, Shijiazhuang, China

Foxtail millet (*Setaria italica*) is nutritious and tolerant to drought and barren conditions, which is an important crop in northern China where there is severe water shortage. However, foxtail millet is very sensitive to herbicide, while its 1000-grain weight is only about 3.0g which makes precision seeding difficult. Thus traditionally seedling thinning and weed control by hand have limited the scale of millet planting. The millet planting area in China has shrunk by 70% in the past 30 years to the current about 1.05 million hm^2.

Millet materials with resistance to Sethoxydim, Imazethapyr and Nicosulfuron which controlled by a single dominant gene, respectively, have obtained by crossing Chinese cultivated millet varieties with green foxtail (*Setaria viridis*) herbicide-resistant mutants coming from Canada since 1993. The above materials have been applied in the following four areas: firstly applied for field seedling thinning and weed control by mixing herbicide-resistant and -susceptible sister lines at certain proportion. Secondly by crossing herbicide-susceptible male-sterile line with herbicide-resistant restoring line, the self-crossed sterile line could be killed when herbicide applying. Thirdly true and false hybrids could be identified when using herbicide-resistant material as male parent crossing with susceptible materials. Fourthly the application of herbicide resistant varieties and hybrids has overcome the obstacle of seedling thinning and weed control by hand in field, which realized large-scale millet production for the first time in China.

W859: Soybean Genomics

Exploring Selective Sweeps Controlling Various Traits in Archaeological Soybean

Sue K Kim[1], Dani Satyawan[1,2], Moon Young Kim[1], HyunJu Jang[1] and **Suk-Ha Lee**[1], (1)Seoul National University, Seoul, South Korea, (2)ICABIOGRAD, Bogor, Indonesia

The history of soybean (*Glycine max*) domestication was explored using genome sequence data obtained from archaeological soybeans dated at 1208 AD. The frequencies of the ancient soybean alleles were calculated in 302 modern soybean accessions comprising wild, landrace, and cultivated soybeans in chromosomal regions that underwent selective sweep for domestication and improvement traits. Most of the favored alleles for domestication traits are present in the ancient soybean chromosomes, although in some segments the frequency of ancient soybean alleles is low in cultivated soybean but common in landraces and wild soybean. This suggests that most of the domestication traits had been selected at 1208 AD, but some modern cultivars had accumulated new alleles in the last 800 years. In selective sweep regions associated with improvement traits, there are several segments where the ancient soybean shows more resemblance to modern cultivars than landraces, suggesting that selection for those segments had been completed 800 years ago. Nevertheless, there are also segments where the ancient soybean is more similar to landraces than modern cultivars, which indicate that the ancient soybean was not selected for the associated improvement trait. Since the selective sweep regions associated with improvement traits often coincide with traits that are important for modern application like oil content, it is understandable that such trait was absent 800 years ago since there was possibly no selection for such trait in that period. Overall, the results indicate a surprisingly advanced selection progress being achieved 800 years ago in a soybean sample obtained in Korea.

W860: Soybean Genomics

Molecular Basis of Soybean Stem Architecture

Jianxin Ma, Purdue University, West Lafayette, IN

Soybean stem growth habit is an important morphological trait, which affect soybean yield potential and adaptability. Recently, we have identified two genes, *Dt1*, and *Dt2*, that control the stem growth habit in soybean, a critical agronomic trait that affects the plant's yield potential and adaptability, and analyzed functional conservation and divergence between these genes and their duplicates. We found that the *Dt1* is the functional counterpart of the *Arabidopsis TFL1* – a floral suppressor gene primarily expressed in shoot apical meristems (SAMs) to prevent their switching from vegetative to reproductive state and thus produces indeterminate stems. The three *Dt1* paralogs generated by the two rounds of WGDs in soybean were functionally diverged from *Dt1* due to their sequence divergence in promoter regions, instead of protein-coding regions. In addition, our data revealed that *Dt2* was a dominant MADS domain factor gene homologous to the *Arabidopsis* floral meristem (FM) identity gene *AP1*. However, unlike *AP1*, whose expression is limited to FMs in which the expression of *TFL1* is repressed, *Dt2* appears to be able to express in SAMs to repress the expression of *Dt1* to promote early conversion of the SAMs into reproductive inflorescences, resulting in semi-determinate stems. We demonstrated that *Dt2* was a recent gain-of-function mutation in its *cis*-regulatory region, which re-shaped the regulatory networks in semi-determinate soybean.

W861: Soybean Genomics

Omics Insights into Isoflavonoid Biosynthesis in Soybean: Functional Specificity of Gene Family Members

Sangeeta Dhaubhadel[1,2], Mehran Dastmalchi[2], Arjun Sukumaran[2] and Caroline Sepiol[2], (1)London Research and Development Centre, Agriculture and Agri-Food Canada, London, ON, Canada, (2)Department of Biology, University of Western Ontario, London, ON, Canada

Soybean (*Glycine max* [L.] Merr) is one of the largest crops grown in Canada, yielding approximately 5.5 million acres, and producing $2.4 billion in profit per year. However, there is a significantly large yield loss due to root and stem rot disease caused by *Phytophthora sojae*. Many strategies have been implemented throughout the years to combat the pathogen such as the use of pesticides and certain agricultural practises. An

alternative approach to this problem is to select a trait naturally found in soybean, such as glyceollin, an isoflavonoid coumpound, production, that can increase innate resistance. Glyceollins act as antimicrobial agents that are synthesized from the isoflavonoid branch of general phenylpropanoid pathway. Isoflavonoids are actors in symbiosis with nitrogen-fixing bacteria, and are also noted for their human health benefits. By transcriptomic analysis, we identified *CHS7* and *CHS8* as genes that are critical for isoflavonoid synthesis in soybean seeds. Further, specific members within the pathway gene families demonstrated tissue- and stress-specific responses suggesting that only selected members of a gene family function during pathogen infection. Additionally, proteomic study identified key players in the isoflavonoid metabolon, the long-postulated model of a subcellular isoflavonoid enzyme complex forming at the surface of the endoplasmic reticulum.

W862: Soybean Genomics

The Role of Gene Function on The Fate of Duplicated Genes in Soybean

Chunming Xu, Kyung Do Kim and Scott A. Jackson, University of Georgia, Athens, GA

Whole genome duplication (WGD) occurred repeatedly during plant evolution. Several models have been proposed to elucidate the evolution of duplicated genes. Most models assume that the duplicated genes evolve randomly and independently. Here, by studying soybean genome which has experienced two rounds of WGD ~13 and 59 MYA, we found that the fate of duplicated genes was highly influenced by their annotated functions. We found the functional categories are unevenly overrepresented in groups of genes with different duplication status. Non-syntenic single-copy genes were significantly enriched in many housekeeping functions, while high-copy (> 4 syntenic copies) genes were highly enriched in transcription regulation functions. There were fewer overrepresented categories in moderate-copy group (2-4 syntenic copies). These facts are consistent with the prediction of scaling laws in the functional content of genomes. We also found that the three gene groups showed dramatically different profiles in expression, DNA methylation and siRNA abundance. By analyzing presence–absence variations (PAVs) among 60 soybean accessions, we determined that 85% of PAV events were found in the non-syntenic group indicating ongoing evolution. Our results suggested that many factors may be involved in the evolution of duplicated genes and that a gene's fate is discriminated by the nature of the gene. We propose that different requirements for complexity in different functional categories by evolutionary selection result in the phenomena of the scaling law of genome functional contents as well as the discriminated evolutionary fates of duplicated genes in different functional categories.

W863: Soybean Genomics

Soybean Transporter Database (SoyTD): Genome-Wide Identification and Exploration of Natural Variants in Soybean Transporter Genes

Gunvant Patil, University of Missouri, COLUMBIA MO, MO

Transporter proteins are membrane channels/pumps which plays a key role in exchange of selective molecules and ions from the external or internal environment. We developed an integrated database for soybean membrane proteins, *SoyTD* that facilitates the interpretation of transporter genes and protein sequence data by integrating features that are available from individual sources. Using publically available prediction programs, putative integral membrane proteins were identified among the 56,044 proteins in the soybean genome (*Wm82.a2*.v*1*). The prediction from the transporter classification database (TCDB) identified approximately 7,541 proteins as transmembrane (TM) candidate proteins. Out of these, 3,306 non-redundant transporter genes containing at least two transmembrane domains were identified and classified according to the transporter classification (TC) system. Genome-wide comparative analysis of transporter genes in 49 plant genomes provided insights into the evolution, expansion and duplication of transporter genes in the land plants. Whole genome resequencing (WGRS) data and tissue specific transcriptome datasets of soybean were integrated to investigate the natural variants and expression profiles associated with transporter(s) of interest. Overall, the SoyTD (http://soykb.org/transporters/) provides a comprehensive interface to soybean researcher to study genetic and molecular function and, will aid as an important resource for soybean improvement.

W864: Speciation Genomics

Genetic Structure Reveals a History of Multiple Independent Origins in the Allopolyploid Weed *Salsola ryanii*

Shana R. Welles, University of Arizona, Tucson, AZ

W865: Speciation Genomics

Asymmetric Origins of a Repeatedly Formed Neo-Allopolyploid Species (*Mimulus peregrinus*)

Joshua R. Puzey, College of William and Mary, Williamsburg, VA

Hybridization between diploids and tetraploids can lead to new allopolyploid species, often via a triploid intermediate. Viable triploids are often produced asymmetrically, with greater success observed for "maternal-excess" crosses where the mother has a higher ploidy than the father. Here we investigate the evolutionary origins of *Mimulus peregrinus*, an allopolyploid thought to be derived from the triploid *M. x robertsii*, to determine whether reproductive asymmetry has shaped the formation of this new species. Using targeted, high-depth sequencing of 1200 genic regions, we confirm the parental origins of this new species from *M. x robertsii*, a sterile triploid hybrid between the two introduced species *M. guttatus* and *M. luteus* that are naturalized and widespread in the United Kingdom. We then used reciprocal crosses between the diploid (*M. guttatus*) and tetraploid (*M. luteus*) progenitors to determine the viability of triploid hybrids resulting from paternal- versus maternal-excess crosses. Organellar sequences obtained from pre-existing genomic data, supplemented with additional genotyping was used to establish the maternal ancestry of multiple *M. peregrinus* and *M. x robertsii* populations. We find strong evidence for multiple and asymmetric origins of *M. peregrinus*, but opposite to the common pattern, with paternal-excess cross significantly more successful than maternal-excess crosses.

W866: Speciation Genomics

Dune Adaptation Facilitates Speciation in Sunflowers

Katherine Ostevik, University of British Columbia, Vancouver, BC, Canada

Sister sunflower taxa are often found in divergent habitats. For example, both *H. neglectus* and an ecotype of *H. petiolaris* inhabit active sand dunes while their closest relative is found on sand sheets. We explore ecological divergence in these two dune systems and assess their progress

towards speciation. We find that larger seeds have evolved in both dune habitats and that the genetic basis of those differences are similar. We also find that assortative mating via conspecific pollen precedence arose surprisingly early between the *H. petiolaris* ecotypes while major chromosomal change is associated with the evolution of *H. neglectus*. Ultimately, understanding the similarities and differences between these systems will help answer the question - how predictable is speciation?

W867: Speciation Genomics

Quantifying Transcriptome-Wide Levels of Divergent Gene Expression Following Speciation By Autopolyploidy in Tolmiea (Saxifragaceae)

Clayton J Visger[1], Gane Ka-Shu Wong[2], Pamela S. Soltis[3] and Douglas E. Soltis[3], (1)University of Florida, Gainsville, FL, (2)Department of Biological Sciences, University of Alberta, Edmonton, AB, AB, Canada, (3)University of Florida, Gainesville, FL

Allopolyploidy has historically been the focal point of researchers interested in the effects of whole-genome duplication. This focus has come at the near exclusion of autopolyploidy and reflects the traditional view of autopolyploidy being evolutionarily unimportant and rare in nature. Recently, the ever-broadening interest in the evolutionary implications of polyploidy has led to the discovery that autopolyploidy is frequent and perhaps a major evolutionary force, particularly in certain clades. Unfortunately, we know little about the ecological, physiological, and genomic implications of speciation by autopolyploidy. The angiosperm genus *Tolmiea* (Saxifragaceae) is an emerging evolutionary system for the study of speciation by autopolyploidy, comprising two species, the autotetraploid *T. menziesii* ($2n = 28$) and its diploid progenitor, *T. diplomenziesii* ($2n = 14$). We present the results of our study of transcriptome-wide divergence in gene expression following autopolyploidy in *Tolmiea.* We also discuss our methods for normalizing expression count data of different ploidal levels through the use of spike-in RNA standards. Using this approach, we report on the differences in total transcriptome size, the proportion of the transcriptome expressed additively versus non-additively (dosage-insensitive), and patterning with respect to ontology. Additionally, we leverage results from our recent and ongoing investigation of ecological and physiological divergence following speciation by autopolyploidy in *Tolmiea* to place the gene expression data in an ecological framework.

W868: Speciation Genomics

So, We Meet Again…Gene Expression in Recurrent Origins of the Allopolyploid Fern *Polypodium hesperium*

Erin M. Sigel, National Museum of Natural History, Smithsonian Institution, Washington, DC

Allopolyploid formation is a common mode of speciation in ferns, with many taxa having formed recurrently from multiple, distinct hybridization events between the same parent species. Each hybridization event marks the union of divergent parental gene copies, or homeologs, and the formation of a novel, independently-derived lineage of an allopolyploid taxon. Little is known about the effects of recurrent origins on genomic composition, phenotypic diversity, and, ultimately, the evolutionary success of allopolyploid ferns. As an initial attempt to address these questions, we have adopted the allotetraploid fern *Polypodium hesperium*, derived from the diploid species *P. glycyrrhiza* and *P. amorphum*, as a natural model system for investigating gene expression patterns between two independently-derived lineages. *Polypodium hesperium* has at least two reciprocally-derived lineages, each with disjunct geographic distributions in western North America. We used Illumina sequencing to construct a *de novo* reference transcriptome and quantify gene expression for three biological replicates of each reciprocal lineage of *P. hesperium* that we grew in common garden conditions. By surveying the gene expression levels for 19000 genes and homeolog-specific expression for 2000 genes, we discovered that gene expression in both reciprocal lineages of *P. hesperium* broadly reflects gene expression in *P. amorphum*—both by mirroring expression levels of *P. amorphum* and preferentially expressing homeologs derived from *P. amorphum*. However, despite similar gene expression patterns across the genome, we recovered substantial variation between the two reciprocal lineages in expression levels and homeolog-expression bias of individual genes. Our results suggest that general “rules” may dictate gene expression patterns in allopolyploids, but recurrent origins impart substantial expression, or phenotypic, variation to an allopolyploid taxon.

W869: Statistical Genomics

Multivariate Genomic Selection in Rice

Zhenyu Jia, University of California Riverside, Riverside, CA

W870: Statistical Genomics

Response Surface Methodology in Genomic Selection

Reka Howard, Iowa State University, Ames, IA

One of the main objectives in plant breeding is to develop sustainable varieties with high yield and desirable characteristics to feed the World. There is an enormous amount of research dedicated to improving plant breeding by developing new methods to increase yield while minimizing cost and the negative effect on the environment. Often the new methods are evaluated using a limited amount of factor combinations (e. g. one or few environments, fixed number of individuals and markers, etc.) due to budget, time or space constraints. Yet, it is important to examine the performance of methods in all possible situations. By focusing on only one or a few factor combinations we may obtain misleading results because model performance can be highly dependent on the attributes of the experiment. Response Surface Methodology (RSM) is an optimization technique that permits estimating the combination of factor level combinations that maximize (or minimize, in some cases) the response without performing the experiment at every possible combination of the variables.

We introduce RSM as a strategy to find the combination of attribute levels that results in accurate predictions for a given GS method, and compare GS methods. We illustrate RSM with an example where we simulate backcross populations with different number of individuals, markers, QTL, and different percentage of epistasis and heritability. The response we optimize is the difference between prediction accuracy using the parametric best linear unbiased prediction method and the nonparametric support vector machine method. This response can be an important diagnostic tool to determine the underlying genetic architecture. The greatest impact on the response is due to the genetic architecture of the population and the heritability of the trait. When epistasis and heritability are highest, the advantage of using the nonparametric support vector versus the parametric best linear unbiased prediction method is greatest.

W871: Statistical Genomics

Optimal Design of Genomic Prediction in Maize Hybrid Breeding

Tingting Guo[1], Xiaoqing Yu[1], Xianran Li[1], Chengsong Zhu[1], Sherry Flint-Garcia[2], Michael D. McMullen[3], James B. Holland[4], Randall J. Wisser[5] and Jianming Yu[1], (1)Department of Agronomy, Iowa State University, Ames, IA, (2)USDA-ARS and University of Missouri, Columbia, MO, Columbia, MO, (3)USDA-ARS, Columbia, MO, (4)North Carolina State University, Raleigh, NC, (5)University of Delaware, Newark, DE

Given a set of *n* parental lines, generating and phenotyping all n^2-*n* reciprocal hybrids are formidable challenges, when $n > 100$. Although genomic prediction is expected to identify promising hybrids with phenotyping some of hybrids in the yield trails, optimal designs are urgently needed to allocate available costs and resources efficiently in commercial breeding. Here we propose an optimal design of genomic prediction based on two criteria: connectedness between training and test populations, and diversity within training population. We test this new strategy in a partial diallel design, comprised of 276 crosses from 24 maize diverse founders in nested association mapping. Our results suggest that this design outperforms other existing designs in terms of prediction accuracy and reliability. We also show how these two criteria affect prediction results separately. Our results prove the advantage of integrating these two criteria in genomic prediction. Furthermore, this research provide a general guideline for the design of genomic prediction project in maize hybrid breeding.

W872: Statistical Genomics

Reconstruction of Genome Ancestry Blocks in Complex Plant Populations

Marco C.A.M. Bink[1], Chaozhi Zheng[1], Julie Ho[2], Christine A. Hackett[3], Johannes Jansen[1], Roeland E. Voorrips[4], Martin P. Boer[1] and Fred E. van Eeuwijk[1], (1)Wageningen UR - Biometris, Wageningen, Netherlands, (2)Forage Genetics International, Nampa, ID, (3)Biomathematics and Statistics Scotland, Dundee, United Kingdom of Great Britain and Northern Ireland, (4)Wageningen UR Plant Breeding, Wageningen, Netherlands

Complex plant populations such as multiparental populations and polyploid families require novel statistical approaches to enable QTL analysis. Here we present two hidden Markov model frameworks, coined RABBIT and TetraOrigin, that allow the reconstruction of genome ancestry blocks in diploid multiparental populations [1] and in tetraploid full-sib families, respectively. The model underlying RABBIT accounts for the joint pattern of recombination breakpoints between two homologous chromosomes and accounts for missing data and allelic genotyping errors in the genotype data of both sampled individuals and founders. Compared to diploids, in tetraploids additional complexity needs to be addressed including the phasing of eight (2*4) parental homologous, double reduction and preferential pairing of chromosomes. We evaluated our methods, and compared them to other methods, by simulation studies and using real data from an *Arabidopsis* MAGIC population and a potato full sib family. We showed that RABBIT is more robust and accurate in reconstructing the genome ancestry blocks, and that TetraOrigin yields parental linkage phasing that is robust to complex chromosome pairing behaviors during meiosis and to erroneous genetic maps, and that the subsequent probabilistic ancestral inference is accurate.

[1] Zheng C, Boer MP, van Eeuwijk FA: Reconstruction of genome ancestry blocks in multiparental populations. *Genetics* 2015, **200**(4):1073-1087.

W873: Statistical Genomics

The Power of Fourier and Wavelet Transforms of Genetic Data in Genotype-phenotype Association Tests under both Monogenic and Polygenic Inheritance

J. Jane Tosh, Shuhua Zhan, Cortland Griswold and **Lewis Lukens**, University of Guelph, Guelph, ON, Canada

Genetic association tests seek to determine if a candidate gene or region contributes to phenotypic variation. Usually, the causative polymorphism within a region is not among the set of genotyped loci. To be detected, the causative locus must be highly correlated with at least one genotyped locus. In these cases, combining genotype information across a region should increase the power to detect causative loci. Nevertheless, a multiple marker approach can increase the degrees of freedom of a test. Approaches involving basis-function transforms (Fourier and wavelet) have been proposed to decrease degrees of freedom. We define a powerful score statistic, and we compare its power to detect associations using untransformed (raw genotype) data and transformed data (both Fourier and wavelet). Our results indicate that over a broad set of genetic architectures, association tests with raw genotype data have more power than tests with transformed data. In scenarios of polygenic inheritance, associations between traits and loci not closely linked to causative loci occur at high frequencies.

W874: Statistical Genomics

A Tutorial of Meta-Analysis for Genome-wide Association Studies

Shizhong Xu, University of California, Riverside, CA

Statistical methods for meta-analysis of genome-wide association studies will be reviewed. Emphasis will be placed on the random model approach to meta-analysis, in which the estimated marker effect of each locus from each study will be treated as a random effect following a normal distribution with an unknown mean (main effect) and an unknown variance. Likelihood ratio and score tests will be used in the study. Rejection of the null hypothesis of zero mean and zero variance indicates a positive detection of the association of the current marker with the trait of interest.

W875: Sugar Beet Workshop

Sugar Beet BeetMap-3, and Steps to Improve the Genome Assembly and Genome Sequence Annotation

Bernd Weisshaar[1], Heinz Himmelbauer[2], Thomas Schmidt[3], Daniela Holtgräwe[4], Andre E. Minoche[5], Juliane C. Dohm[2], Ralf Stracke[1], Falk Zakrzewski[3], Roza Parol-Kryger[1], Kai B. Stadermann[1], Jessica Schneider[1], Thomas Rosleff Soerensen[1], Thomas Kraft[6] and Britta Schulz[7], (1)Bielefeld University, Department of Biology & CeBiTec, Bielefeld, Germany, (2)Universität für Bodenkultur, Department of Biotechnology, Vienna, Austria, (3)Dresden University of Technology, Department of Biology,

Dresden, Germany, (4)Department of Biology & CeBiTec, Bielefeld, Germany, (5)Center for Genomic Regulation, Barcelona, Spain, (6)Syngenta Seeds AB, Landskrona, Sweden, (7)KWS SAAT SE, Einbeck, Germany
The sugar beet genome sequence was brought to publication quality within AnnoBeet and was finally published in January 2014. Since then, the AnnoBeet consortium has focused on further improvements of the genome assembly and its annotation on the one hand, and on steps towards using genomics resources for answering biological questions on the other hand.
By using a combination of existing resources for SNP calling, a set of new SNP markers was developed. The data used were Sanger EST sequences from KWS2320, representing P1 of our main mapping population, and a 454 cDNA data set from P2 of that population. For a subset of the about 1,600 loci with good SNP predictions, markers were developed and mapped. In addition, terminal chromosomal markers from Paesold et al. (2012) were integrated. Finally, the sugar beet genetic map was extended by 307 markers, resulting in BeetMap-3 comprising 1,141.4 cM.
Also, we improved the assignment of unanchored contigs to pseudochromosomes by 'genotyping by sequencing' using low coverage NGS data from F2 plants of the mapping population, and in addition offspring of an additional mapping population with two parents unrelated to P1 and P2. As a result, the assembly was improved and the number of unassigned or unanchored contigs and scaffolds was reduced. Additional measures to further increase the quality of the assembly are under way, including the evaluation of PacBio reads as well as Moleculo data for joining contigs.
Based on SMRT (PacBio) cDNA sequencing reads, deep Illumina RNA-seq data and manually annotated gene models, an improved gene set for sugar beet was created using the AUGUSTUS software. The new BeetSet-2 covers 26,923 protein coding genes that are predicted with high reliability and are supported by evidence. In addition, there are about 13,000 genes that are predicted with the improved AUGUSTUS parameters specific for sugar beet but lack experimental evidence.
Finally, we are using the genomics data to investigate specific gene families and try to validate the functional annotation of genes that has been assigned *in silico*. A total of 70 *R2R3-MYB* genes as well as genes encoding three other classes of MYB proteins containing multiple MYB repeats were identified and characterized with respect to amino acid sequence, structure and chromosomal organization. Functional classification of the gene family led to the identification of a sugar beet-specific clade with an atypical amino acid composition in the R3 domain that were predicted to encode betalain regulators. The functional classification was further verified by experimental confirmation of the prediction that the R2R3-MYB gene *Bv_iogq* encodes a flavonol regulator which was designated BvMYB12.

W876: Sugar Beet Workshop

Molecular Characterization of Wild *Beta* Populations in the Imperial Valley, California

Kelley Richardson, USDA-ARS, Crop Improvement and Protection Research Unit, Salinas, CA
Populations of wild *Beta* species exist as weeds in commercial sugar beet fields in the Imperial Valley, California. Significant losses to sugar yield and quality result if these wild plants are not removed. In cases of extreme infestation, fields are abandoned without harvest. No selective chemicals are available to differentiate conventional sugar beet from wild relatives and hand removal is labor intensive and expensive. Planting sugar beet varieties with tolerance to glyphosate is a potential solution for infested fields, but risk of gene flow to adjacent wild relatives must be determined. Previous research identified these populations as either *Beta vulgaris* subspecies *maritima* or *Beta macrocarpa*. This distinction is critical because *B. v*. ssp. *maritima* will readily cross hybridize with cultivated sugar beet while *B. macrocarpa* rarely will. In April 2011, whole plants, mature seed, and leaf tissue for DNA extraction were collected from wild plants in 25 infested sugar beet fields throughout the Imperial Valley. Morphology of plants from collected seed grown in non-competitive conditions assigned taxonomy of these populations to *Beta macrocarpa*. In this study, we used molecular tools with the objective to determine genetic similarities and differences between and within wild *Beta* populations. DNA was extracted from collected leaf tissue and DNA of known *B. v*. ssp. *maritima* and *B. macrocarpa* accessions was included for comparison. Six simple sequence repeat (SSR) molecular markers were run and fluorophore-assisted fragment analysis assigned sizes to resulting PCR products. Marker data was evaluated for population structure. The DK statistic divided the wild *Beta* populations into two subgroups suggesting gene flow occurred either amongst wild beet populations or between cultivated and wild beets at some point in their history. Phylogenetic analyses are underway to further clarify these relationships.

W877: Sugar Beet Workshop

Digital PCR for Genotyping of SNP_BvBTC1 in DNA Bulks of Sugar Beet

Piergiorgio Stevanato, University of Padova, Italy, Legnaro (Padova), Italy
Discovery and validation of SNP markers linked to the annual habit in Beta vulgaris is essential for breeding of sugar beet. Pin et al. 2012 (Curr. Biol. 22:1095-1101) found a SNP variant in exon 8 (nucleotide position 79) of the gene BvBTC1 (BOLTING TIME CONTROL 1) that helps to discriminate between annual and biennial plants. A TaqManTM assay (SNP_BvBCT1) was designed to detect this SNP variant. SNP_BvBTC1 could be genotyped to differentiate annual from biennial flowering plants: CC genotypes are mainly present in biennial beets, while AA and CA genotypes are typically found in annual beets. However, screening many populations for SNP polymorphisms is time-consuming and expensive when performed as individual genotyping. An option to reduce the cost of SNP_BvBTC1 screening for removal of off-type plants is to use DNA bulks, where plant DNA is pooled by group and genotyping is performed on the bulk rather than on individual plants. In this study, we compared the accuracy of the SNP_BvBTC1 genotyping by qPCR, HRM and digital PCR (dPCR) in two DNA bulks constituted by pooling 90 biennial + 10 annual plants (B1) and 99 biennial + 1 annual plants (B2) to simulate high or low contamination with wild beets, respectively. qPCR could not detect the minor allele in either the B1 or B2 pool; HRM detected the allele A only at moderate frequencies (10%), in the B1 pool. dPCR, on the contrary, was able to detect the allele A in both pools. This study provides evidence that dPCR is suitable for the quantitation of SNP_BvBTC1 from pooled DNA samples in sugar beet.

W878: Sugar Beet Workshop

Overexpression of CRK8 Gene in Sugar Beet line Resistant to Beet Curly Top disease

Imad Eujayl, USDA-ARS, Northwest Irrigation and Soils Res. Lab., Kimberly, ID and Carl Strausbaugh, USDA-ARS-NWISRL, Kimberly, ID

Resistance to *Beet curly top virus* (BCTV) is an essential trait for all cultivars grown in western USA. Currently insecticides are used to compliment low and moderate levels of resistance in commercial varieties. To identify genes regulating resistance to BCT, differential gene expression was studied *via* RNA-sequencing using an exceptionally resistant doubled haploid line (KDH13-PI663862) compared to a susceptible breeding line (K19-19). KDH13 was subjected to 7 treatments: infested with non-infectious leafhoppers, infected with leafhoppers population carrying a single, two, or three strains, and control healthy leaf. The susceptible line was infected with the three strains. Leaf samples were sequenced in a HiSeq2500 and data was analyzed using TopHat and Cufflinks software. All sequences were aligned to the RefBeet-1.2. Based on 28 pair-wise comparisons, differentially expressed genes were determined with a cuttoff of false-discovery-rate; FDR<0.05 and a LogFC (expression fold change) >2.0 or <-2.0. In this study we are focusing on the intricate patterns of the expression levels of the protein-coding gene CRK8 (cysteine-rich receptor-like protein kinase), induced by beet leafhopper infestation and BCTV. There are 171 members of the CRK8 (AT4G23160) that were differentially expressed in comparisons between healthy, infested, and infected leaf of KDH13 and K19-19. In KDH13 a single member of the CRK8 showed a consistent overexpression after leafhopper infestation or infection with the three strains at LogFC of 2.3 and FDR $<1.0^{-4}$. qPCR analysis was conducted to support these results for 4 CRK8 members using tissue from the same plants used for the initial RNA sequencing. The qPCR results confirm overexpression of two of the CRK8 members in the resistant line.

W879: Sugar Beet Workshop

Metabolomic Profiling to Characterize the Defense Response of Sugar Beet during *Rhizoctonia solani* Interactions

Kimberly Webb, USDA-ARS- Sugar Beet Unit, Ft. Collins, CO

Sugar beet can be significantly impacted by Rhizoctonia crown and root rot caused by *Rhizoctonia solani* AG 2-2 IIIB. The molecular processes that mediate compatible and incompatible sugar beet interactions with *R. solani* are largely unknown and identifying the metabolites associated with *R. solani* infection may provide evidence for important biological pathways involved with resistance. The metabolic changes that occurred during susceptible and resistant *R. solani* interactions were compared with mock inoculated treatments and characterized using a non-targeted metabolomics workflow spanning primary and secondary metabolism products. Metabolites were extracted from infected and healthy, root and leaf tissue at multiple time points to best reflect the *R. solani* infection process and analyzed using both UHPL-MS and GC-MS. Statistical interrogation of the datasets revealed clear distinction between tissue type and genotype, and more subtle changes in response to inoculation that was dependent on genotype. UHPLC-MS was more sensitive than GC-MS, and therefore more secondary metabolites including several phytolexins, terpenes, and alkaloids, were identified. Several biochemical pathways appear to be involved during susceptible and resistant interactions with *R. solani*, and indicate a complex role of primary and secondary metabolites in sugar beet during fungal interactions.

W880: Sugar Beet Workshop

The Genome of *Chenopodium quinoa*

David E Jarvis, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Chenopodium quinoa (quinoa) is a relative of sugar beet and spinach and is native to the Andean region of South America, where it has adapted to tolerate extreme abiotic stresses, including salinity, drought, heat and frost. Although quinoa has been a staple food source in the region for thousands of years, it has only recently gained international attention due to the high protein content and balanced amino acid composition of its seeds. Short-read sequencing technology previously limited the creation of a gold-standard reference assembly in this allotetraploid (AABB) species. Recent advances in Pacific Biosciences long-read technology have allowed the sequencing of quinoa to approximately 62X coverage, with the aim to understand the genomic features underlying quinoa's nutritional properties and abiotic stress tolerance. The draft 1.2 Gb assembly is made up of 4971 contigs, with 50 % of the genome assembled in contigs greater than 871 Mb. In addition, 245 Gb (approximately 200X coverage) of BioNano optical map data was generated with the Irys technology to further improve the assembly. The genome annotation was guided by 45 different RNAseq libraries from multiple tissues and treatments, with the goal to generate a gold-standard transcriptome. In addition, the genomes of the related diploid species *C. pallidicaule* (AA) and *C. suecicum* (BB) and the related tetraploid species *C. hircinum* and *C. berlandieri* have been sequenced to help resolve the evolutionary relationships between quinoa and its relatives. These resources will serve as valuable tools in future genetic improvement of quinoa to increase food security in the face of global climate change and a growing world population.

W881: Sugar Beet Workshop

***Beta vulgaris* Crop Types: Genomic Signatures of Selection (GSS) Using Next Generation Sequencing of Pooled Samples**

Paul Galewski, Michigan State University, East Lansing, MI and J. Mitchell McGrath, Sugar beet and Beans Research Unit, East Lansing, MI

Beta vulgaris crop types represent highly diverged populations with distinct phenotypes resulting from long-term selection. Differential end use in the crop types includes: leaf quality (chard/leaf beet), root enlargement and biomass, (table beet, fodder beet, sugar beet), and secondary metabolite accumulation (sugar beet, table beet). Many of the defining features of the crop types are highly reproducible across environments suggesting that selection directly affected many genes across the genome. Variation in allele frequency among crop type populations provides a means to detect GSS and thus identify genic regions influencing some economically important traits. We sequenced six populations representing four of the *B. vulgaris* crop types by pooling 25 individuals per population, created PCR-free libraries, and sequenced each population to ~100X depth. Reads aligned to the RefBeet1.1 genome assembly were used to evaluate nucleotide diversity ($\boldsymbol{\pi}$), inbreeding coefficients (F), and population differentiation (Fst). These measures were plotted against the RefBeet1.1 chromosomes and regions under positive selection were evaluated across each crop type genome and functional annotations for genes underlying the regions of interest were extracted. Understanding the function and ontology of genes under selection and their role in producing crop type variation is the primary impetus for this work.

W882: Sugar Cane (ICSB)

Biotechnology for Sugarcane Improvement in CSIRO Australia

Anne L. Rae[1], Karen S. Aitken[1], Graham D. Bonnett[1], Rosanne E. Casu[1], Phillip Jackson[2] and **Lynne McIntyre**[1], (1)CSIRO Agriculture, St Lucia, Australia, (2)CSIRO Agriculture, Aitkenvale, Australia

Sugarcane is a significant industry in the north-eastern region of Australia, generating approximately $2B in revenue annually. Although conventional breeding and selection has been very successful in maintaining productivity, yields have not increased at the expected rate when compared to other crops. CSIRO has worked with sugarcane breeders and researchers over recent years to develop novel approaches for breeding and selection, based on research into the genetics of this complex polyploid. The development of high density genetic maps in Australian cultivars has provided valuable insights into the genetic structure of sugarcane as well as the basis for the assessment of both genetic variation and genetic associations in sugarcane. Genetic variation in a range of traits has been assessed, including sugar and fibre content, water use efficiency, plant architecture and disease resistance. Genetic associations have been determined through the construction of high density genetic maps and transcriptomic analyses. More recently, we have contributed to the development of the genome sequence of sugarcane and its relatives to better dissect gene function and to develop SNPs. A 40K SNP chip is currently being validated for use in selection for disease resistance in late-stage cultivar development and for genome wide selection for parental improvement. Transgenic approaches have also been tested and the group has delivered data to enable regulatory assessment of GM varieties by measuring the likelihood of gene flow from commercial sugarcane via seed production and establishment.

W883: Sugar Cane (ICSB)

Introgression of a Large Effect QTL for Smut Resistance Inherited from *S. spontaneum*

Karen S. Aitken[1], Shamsul Bhuiyan[2], Barry Croft[2], Nathalie Piperidis[3], George Piperidis[3], Phillip Jackson[4] and Meredith McNeil[1], (1)CSIRO Agriculture, St Lucia, Australia, (2)Sugar Research Australia, Woodford, Australia, (3)Sugar Research Australia, Te Kowai, Australia, (4)CSIRO Agriculture, Aitkenvale, Australia

Smut caused by *Sporisorium scitamineum* is a major sugarcane disease that can cause considerable yield losses. Sugarcane resistance to smut has been demonstrated to be heritable although the genetic determinants of this resistance are unknown. Studies involving modern cultivars have shown inheritance to be highly quantitative with many small effect QTL (9% variation explained) contributing to resistance. Introgression from wild species has long been of interest to sugarcane breeders to increase the genetic diversity of modern cultivars as only a small number of original progenitor clones are believed to have contributed to the ancestry of sugarcane. *S. Spontaneum* is widely recognised as imparting a range of important traits to modern sugarcane cultivars such as adaptation to environmental stresses, ratooning performance and resistance to diseases. An introgression program was started in Australia with the introduction of 43 biparental crosses between sugarcane and *Saccharum spontaneum*. These populations were evaluated for performance in field trials and for disease resistance. Progeny from a cross between a BC1 *S. spontaneum* clone and a sugarcane cultivar were found to segregate for resistance to smut. AFLP, SSR and DArT markers have been used to map this population and QTL analysis has identified a major effect that varies from 36% to 45% variation explained depending on year. This population is being screened across a sugarcane SNP chip to improve the resolution of the genetic map. Further crossing is being carried out to introgress this region into commercial cultivars.

W884: Sugar Cane (ICSB)

Can Cytogenetic and PCR Markers Assist Selection of High Value *Erianthus*-Derived Sugarcane Clones?

Nathalie Piperidis[1], Karen S. Aitken[2] and George Piperidis[1], (1)Sugar Research Australia, Te Kowai, Australia, (2)CSIRO Agriculture, St Lucia, Australia

Intergeneric hybrids between *Saccharum* and *Erianthus* are the newest exotic addition to the Australian sugarcane breeding program. The wild species *Erianthus arundinaceus*, a close relative of sugarcane could be an important candidate to enlarge the gene pool of the Australian parent populations. *E. arundinaceus* is reported to contain numerous traits of agronomic importance including pest and disease resistance and tolerance to drought and water-logging conditions. Based on genomic *in situ* hybridisation (GISH) results on the *Erianthus* introgression clones, we have developed a molecular method to allow selection of hybrids at an early stage that have incorporated sought-after traits. So far, five generations of fertile hybrids from intergeneric crosses between *S. officinarum, E. arundinaceus* and modern sugarcane cultivars have been produced. The chromosome composition of F1, BC1, BC2, BC3 and BC4 hybrids was studied by GISH and revealed a reduction of *Erianthus* chromosomes as well as a loss of chromosomes from the BC1 generation. We also revealed the formation of recombinant chromosomes between both genera. Individuals with low number of *Erianthus* chromosomes or recombined chromosomes could be an important addition to the breeding program if these chromosomes are inherited stably during crossing. The stability of recombined chromosomes is currently under investigation by studying the transmission of these recombined chromosomes in further crosses. Nevertheless, as the number of *Erianthus* chromosomes in the BC3 (and BC4) are lower or equivalent to the basic number of *E. arundinaceus* (x=10), we are aiming to develop a simple method based on PCR molecular markers which will allow the identification of individual *Erianthus* chromosomes in the backcross hybrids. If important traits, such as nematode and/or pachymetra resistance can be linked to *Erianthus* specific chromosomes, then this method could become a valuable tool for sugarcane breeders as an effective selection screening method.

W885: Sugar Cane (ICSB)

Advances in the Utilization of Next Generation Sequencing Data in the Colombian Sugarcane Breeding Program

John J. Riascos, Colombian Sugarcane Research Center (Cenicaña), Cali, Colombia

Sugarcane is a polyploid/aneuploid crop with a chromosome number ranging between 2n = 80-130. This genomic complexity has made difficult the implementation of breeding strategies based on molecular markers. Importantly, the recent advances in next generation sequencing technologies (NGS) has increased the number of markers that can be used in genotyping studies and thus offer new possibilities in breeding programs.

At CENICAÑA, a diversity panel of 220 sugarcane genotypes, 130 representing the phenotypic diversity of CENICAÑA's germplasm collection and 90 of breeding and economic importance for the center, was sequenced by means of GBS and RAD. Independently, each dataset was analyzed to identify high quality SNPs, determine the ratio of homozygous:heterozygous SNPs within a genotype, calculate genetic distances among individuals, define populations structure and explore phenotype-genotype, GWAS, associations.

Among our most relevant results we found that phylogenetic trees derived from GBS and RAD were not identical (groups coincided in 60%-70% of the cases), despite the high number of SNP markers used (>5,000). The ratio of homozygous:heterozygous SNPs always favored

homozygosity, which was higher in RAD compared to GBS. These data combined with genetic distances indexes revealed parental genotypes with potential for crossings aiming to increase overall genetic diversity. Population structure analyses detected four subpopulations, which is in agreement with previous studies based on microsatellites. RAD sequencing performed better than GBS in detecting these subpopulations. This information was used in a GWAS approach were preliminary, significant, SNP-alleles were found for traits such as brix, reducing sugars and sucrose production.

W886: Sugar Cane (ICSB)
Coffee break
Nathalie Piperidis, Sugar Research Australia, Te Kowai, Australia

W887: Sugar Cane (ICSB)
Using Droplet Digital PCR (ddPCR) to Detect Copy Number Variation in Sugarcane, a High-Level Polyploid
Per Hilding McCord, USDA Agricultural Research Service, Canal Point, FL
Copy number variation (CNV) generally refers to duplications or deletions of sections of DNA, including genes. It has been implicated in a number of important diseases in humans. In polyploid plant species, CNV can also take the form of multiple copies of an allele at a given locus. This variation in dosage has been shown to alter gene expression and phenotype. As copy number rises, it becomes increasingly challenging to quantify dosage changes. A new technique, known as droplet digital PCR (ddPCR), has been used for measuring CNV. Digital PCR partitions a PCR reaction into many thousands to millions of discrete reactions (emulsified droplets in the case of ddPCR). This partitioning has the effect of diluting the target DNA, such that some droplets contain zero copies of the target, while others contain one or more copies. Post-PCR, this results in a binary (positive or negative) outcome, hence the term digital. Partitioning the reaction into a large number of discrete tests also increases sensitivity and dynamic range. In this article, ddPCR was employed to survey sugarcane (*Saccharum* spp.) germplasm for dosage variation for the *Bru*1 locus, which imparts resistance to the fungal disease known as brown rust. The breeding nursery at the USDA-ARS Sugarcane Field Station in Canal Point, FL was surveyed for the presence of *Bru*1 using standard methods. Those clones which were *Bru*1-positive were then analyzed via ddPCR. Out of 80 *Bru*1-positive clones, the majority (60) were simplex or single copy. Eighteen clones were duplex for *Bru*1, and there were two clones that were triplex. The progeny of several test-crosses supported the results of ddPCR; a simplex parent produced approximately 50% positive progeny, while parents that were duplex produced more than 80%. Further verification was done via quantitative PCR. Results generally agreed with ddPCR, in that three groups could be distinguished, but the groups were relative, and not quantitative. This analysis of dosage variation has important implications for both breeding and molecular biology research. Clones with more than one copy of *Bru*1 (or any other gene of interest) will pass on the gene at much higher frequencies than close that are simplex. In addition, the identification of clones with varying gene dosages allows for formal testing of dosage effects. This is the first report of ddPCR being used to measure copy number/dosage variation in a high-level polyploid plant species.

W888: Sugar Cane (ICSB)
Sugar in Small Bits: Epigenetic Regulation in the Saccharum Complex
Paulo Ferreira, IBqM/Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
Modern sugarcane cultivars are hybrids of *Saccharum officinarum* and *Saccharum spontaneum* species. Hybrids have complex polyploidy genomes that difficult the studies of structure and organization of the genome. To understand role of the small RNA regulation in species of the *Saccharum* complex, small RNA and degradome libraries were prepared from RNA obtained from leaves of *S. officinarum, S. spontaneum, S. sinensis* and *S. robustum*, as well as from two commercial hybrids, RB867515 and SP701143. Deep sequencing identified a large set of conserved miRNA and novel miRNA candidates. Annotation of the confirmed degraded mRNAs identified most of the canonical sequences found in plants. In addition, novel targets, not previously described in plants, were also identified. We used the miTRATA pipeline with a subset of 11 conserved microRNAs to evaluate microRNA turn over in the six genotypes. The results indicate that the rate of turnover is dependent of both microRNA sequence and genotype. Together, our data provides insights on the overall composition of sRNAs in wild *Saccharum* species and hybrids, and also improve the understanding of sRNA-triggered regulation in the context of hybridization in sugarcane.

W889: Sugar Cane Sequencing Initiative
The Genome Sequence of a Sugarcane Hybrid: Exploring Genes, Alleles and Transcripts
Carolina Gimiliani Lembke[1], Pablo Andrade[2], Hayan Lee[3], Gabriel Rodrigues Alves Margarido[4], Milton Y. Nishiyama Junior[5], Savio S. Ferreira[1], Ravi Pandya[3], Jonas W. Gaiarsa[6], Michael Schatz[7], David Heckerman[8], Robert Davidson[3], Marie-Anne Van Sluys[9] and Glaucia M. Souza[10], (1)Institute of Chemistry - University of São Paulo, São Paulo, Brazil, (2)Institute of Chemistry - University of São Paulo, SAO PAULO, Brazil, (3)Microsoft Research, Redmond, WA, (4)USP / ESALQ, Piracicaba, SP, Brazil, (5)University of São Paulo, São Paulo, Brazil, (6)GaTE Lab - University of São Paulo, Sao Paolo, Brazil, (7)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (8)Microsoft, Los Angeles, CA, (9)Universidade de São Paulo, Sao Paulo, Brazil, (10)Universidade de São Paulo (USP), São Paulo, Brazil
The increase in bioenergy efficiency and the transition to a biobased economy depends on the improvement of biomass production, such as sugarcane. This implies in the integration and deepening of the knowledge on agronomic practices, biochemistry and genetics. We are conducting the genome sequencing of a sugarcane cultivar and integrating it with transcriptome data. The challenges in sequencing the sugarcane genome relies in the assembly and analysis of a highly complex genome that is polyploid and aneuploid. To sequence sugarcane genome we combined BAC-by-BAC and shotgun approaches and different sequencing platforms. The development of Illumina Moleculo synthetic long-read sequencing technology showed to be a promising technology to sequence the sugarcane genome. We sequenced 26 libraries and obtained a total of 21.8 Gb of sequencing data, which approximately corresponds to 2.18 X coverage of the 10 Gb polyploid sugarcane genome. The Celera Assembler was modified for usage with longer reads. The assembly yielded 450,609 contigs in 4.26 Gb, with contig length average of 9,452 bp (min, 1,500; max 468,011). For the identification of putative sugarcane alleles we first selected single copy sorghum genes (6,761) and aligned them to the predicted sugarcane coding sequences. The number of copies varied between 1 and 15. With pairwise alignments between the

putative alleles we could verify the distance between them. Aligning cDNA reads from a full-length enriched cDNA libraries from *S. officinarum*, *S. spontaneum* and SP803280, we detected allele expression. For the 10 most expressed GO categories, we evaluated the percentage of alleles expressed.

W890: Sugar Cane Sequencing Initiative
Sugarcane Moleculo Contigs into Chromosome Cluster Clouds
Marie-Anne Van Sluys, Universidade de São Paulo, Sao Paulo, Brazil

W891: Sugar Cane Sequencing Initiative
The Sugarcane Chloroplast Genome: A Next Generation Sequencing Perspective
Nam Van Hoang[1], Agnelo Furtado[2], Richard B. McQualter[1] and Robert J. Henry[2], (1)University of Queensland QAAFI, Brisbane, Australia, (2)University of Queensland/QAAFI, Brisbane, Australia
The sugarcane chloroplast genome sequence and heteroplasmy in sugarcane chloroplast was re-investigated using next generation sequencing. NGS data from total DNA of sugarcane (*Saccharum* spp. hybrids) cultivar Q155 was used for chloroplast genome assembly, employing a combined approach of *de novo* assembly and reference-based mapping. It was found that the chloroplast genome of sugarcane cultivar Q155 is likely to be the chloroplast genome sequence for cultivated sugarcane, as it is the consensus of sequences reported for other two cultivars, NCo310 and SP80-3280. Our finding is consistent within the recent narrow origin of sugarcane in which all modern sugarcane hybrid cultivars were derived from a few selected clones after a few generations within a short-time divergence. We carried out further analysis, using samples with different proportions of nuclear, mitochondrial and chloroplast DNA, to test the potential to distinguish genuine chloroplast heteroplasmy from apparent heteroplasmy due to chloroplast homologues inserted in other cellular genomes. It is revealed that the reported variants were present only at frequencies that could be attributed to non-chloroplast sequences from mitochondria and nucleus. There is no positive evidence from NGS data for heteroplasmy in the sugarcane chloroplast genome. This might also indicate that plant chloroplast genomes do not display heteroplasmy.

W892: Sugar Cane Sequencing Initiative
Long Read Sequencing Technology to Solve Complex Genomic Regions Assembly in Plants
Helene Berges[1], Arnaud Bellec[2], Audrey Courtial[2], Nathalie Rodde[2], Stephane Cauet[2], Sonia Vautrin[2], Gensériс Beydon[2], William Marande[2] and Yves Barriere[3], (1)Plant Genomic Center - INRA Toulouse, Castanet-Tolosan, France, (2)INRA - CNRGV, Castanet Tolosan, France, (3)INRA, Lusignan, France
Numerous whole genome sequencing projects already achieved or ongoing have highlighted the fact that obtaining a high quality genome sequence is necessary to address comparative genomics questions such as structural variations among genotypes and gain or loss of specific function. Despite the spectacular progress that has been done regarding sequencing technologies, accurate and reliable data are still challenging, at the whole genome scale but also when targeting specific genomic regions. These issues are even more noticeable for complex plant genomes. Most plant genomes are known to be particularly challenging due to their size, high density of repetitive elements and various levels of ploidy. To overcome these issues, we have developed a strategy in order to reduce the genome complexity by using the large insert BAC libraries combined with next generation sequencing technologies. We have compared two different technologies (Roche-454 and Pacific Biosciences PacBio RS II) to sequence pools of BAC clones in order to obtain the best quality sequence. We targeted nine BAC clones from different species (maize, wheat, strawberry, barley, sugarcane and sunflower) known to be complex in terms of sequence assembly. We sequenced the pools of the nine BAC clones with both technologies. We have compared results of assembly and highlighted differences due to the sequencing technologies used. We demonstrated that the long reads obtained with the PacBio RS II technology enables to obtain a better and more reliable assembly notably by preventing errors due to duplicated or repetitive sequences in the same region.

W893: Sugar Cane Sequencing Initiative
Genomic Selection on a Sugarcane Breeding Population Using High-Throughput Sequence Capture Genotyping
Marcio Resende[1], Leandro G Neves[1], Ivone de Bem Oliveira[2], Clistiane dos Anjos Mendes[2], Ludmila F. Bandeira[2] and Alexandre S. G. Coelho[2], (1)RAPiD Genomics LLC, Gainesville, FL, (2)Universidade Federal de Goias, Goiania, Brazil

W894: Sugar Cane Sequencing Initiative
Sugarcane is an Ancient Allopolyploid
Andrew H. Paterson, Plant Genome Mapping Laboratory, University of Georgia, Athens, GA

W895: Sweet Potato and Yam Genomics
Advances in Sweetpotato and Yam Genomics
Jim Lorenzen, BMGF, Seattle, WA
A changing world with ever-increasing requirements for food will require acceleration of productivity increases, especially in Sub-Saharan Africa. A key element of this will be breeding for higher productivity. For crops with relatively short breeding histories such as cassava, sweetpotato, and yam, presumably significant yield increases could be achieved in relatively few breeding cycles. This could be further accelerated by modern tools of breeding, including molecular tools, of which an anchored reference genome sequence is extremely valuable. While the list of crops with available reference genome sequences is pushing over 50, sweetpotato and yam are not listed among them (as of Oct. 2015). However, sequencing efforts for sweetpotato and both main species of *Dioscorea* food yams in Asia, the USA, and Europe are delivering quality assemblies, and high-density sequencing-based genotyping has started for these species. This is an exciting time for genomic tools for sweetpotato and yam, and the workshop should deliver a series of interesting and informative talks that bring many more scientists up to date on recent developments for these important root and tuber crops.

W896: Sweet Potato and Yam Genomics

The Development of Breeder-Friendly Genomic Tools for Sub-Saharan Sweetpotato Variety Development - What's needed and how do we get there?

Craig Yencho, North Carolina State University, Raleigh, NC

Sweetpotato is a widely recognized food security and cash crop with highly recognized potential to alleviate hunger, vitamin A deficiency, and poverty in Sub-Saharan Africa (SSA). It is also a crop predominantly grown in small plot holdings by poor women farmers across SSA. The Genomic Tools for Sweetpotato (GT4SP) Improvement Project is working to develop modern genomic, genetic, and bioinformatics tools to facilitate sweetpotato improvement. Our goal is to developing a set of "next generation" breeder tools for African sweetpotato breeders. In this talk I will describe the GT4SP project, discuss what we think is needed to achieve our goal and transfer improved varieties to smallholder farmers across SSA.

W897: Sweet Potato and Yam Genomics

Genome Sequencing and Annotation of Sweetpotato Wild Progenitors

Zhangjun Fei, Cornell University, Ithaca, NY

Sweetpotato, *Ipomoea batatas*, is among the most important food crops in the world and an extremely important food crop for subsistence farmers in sub-Saharan Africa. Despite its significant importance, currently the genomic resources for sweetpotato are very limited. Sweetpotato is an allo-auto-hexaploid (2n=6x=90) with two hypothesized, non-homologous genomes. The polyploid and heterozygous nature of the sweetpotato genome makes genetic and functional analyses extremely challenging. To facilitate research and breeding in this important species, we have assembled the genomes of *I. trifida* and *I. triloba*, two diploid wild progenitors of sweetpotato, using the Illumina and PacBio sequence reads. Both assembled genomes were ~460 Mb, ~87% and 90% of the *I. trifida* and *I. triloba* genomes with estimated sizes of 530 Mb and 510 Mb, respectively. The N50 sizes of the assembled scaffolds are 1.48 Mb and 6.46 Mb for *I. trifida* and *I. triloba*, respectively. High throughput transcriptome sequences have been generated from various different tissues and are being used to facilitate genome annotation. High-density genetic maps and BioNano physical genome maps are also being constructed for anchoring and ordering assembled genome scaffolds. The high-quality *I. trifida* and *I. triloba* genomes will serve as valuable references for extensive QTL mapping, marker identification and comparative genomics of sweetpotato.

W898: Sweet Potato and Yam Genomics

Transcriptome Sequencing of the Sweet Potato Progenitor (*Ipomoea trifida* (H.B.K.) G. Don.) and Discovery of Drought Resistance Genes

Qinghe Cao, Sweetpotato Research Institute, Chinese Academy of Agricultural Sciences, Xuzhou, China

Ipomoea trifida (H.B.K.) G. Don. is the closest wild relative of cultivated sweet potato (*I. batatas*). The diploid *I. trifida* plays an important role in sweet potato breeding and construction of transgenic systems. It also can be used to discover functional genes, particularly stress resistance genes lost during the domestication of sweet potato. However, the transcriptome information of *I. trifida* is limited, compared to *I. batatas* data. Using high-throughput Illumina RNA-seq technology, a total of 66,329,578 paired-end 101 bp reads were cleaned and assembled *de novo* to produce 90,864 *I. trifida* transcripts. Based on sequence similarity, 69,541, 39,236, 12,509, 2,848 and 2,766 transcripts were annotated using homologous proteins from NCBI NR database, GO terms, KEGG pathways, known transcription factors and protein kinases, respectively. Great differences were found between the *I. trifida*. and *I. batatas* transcriptomes.

This is the first report of a comprehensive transcriptome for diploid *I. trifida*, with gene expression information for root, leaf, stem and flower tissues. The *I. trifida* transcriptome sequences enrich the gene resources for sweet potato molecular research and breeding. Moreover, we demonstrated that these sequences could be used to design SSR markers and for gene discovery in *I. trifida*. In particular, we found a potential drought resistance gene ItWRKY1 from *I. trifida* and validated its function using the transgenic tobacco system.

W899: Sweet Potato and Yam Genomics

Whole Genome Sequencing of White Guinea Yam (*Dioscorea rotundata*): Towards Generating Genetic and Genomic Tools for Improvement of an African 'Orphan' Crop

Muluneh Tamiru Oli[1], Satoshi Natsume[1], Hiroki Takagi[1], Kentaro Yoshida[2], Hiroki Yaegashi[1], Aiko Uemura[1], Kaori Oikawa[1], Naoya Uraski[3], Hideo Matsumura[4], Pachakkil Babil[5], Shinsuke Yamanaka[6], Ryo Matsumoto[6], Satoru Muranaka[6], Gezahegn Girma[7], Antonio Lopez-Montes[7], Melaku Gedil[7], Ranjana Bhattacharjee[7], Michael Abberton[7], P. Lava Kumar[7], Ismail Rabbi[8], Guenter Kahl[9], Hiroko Takagi[6], Robert Asiedu[7] and Ryohei Terauchi[1], (1)Iwate Biotechnology Research Center, Kitakami, Japan, (2)Kobe University, Kobe, Japan, (3)Okinawa Prefectural Agricultural Research Center, Itoman, Japan, (4)Shinshu University, Ueda, Japan, (5)Tokyo University of Agriculture, Tokyo, Japan, (6)Japan International Research Center for Agricultural Sciences, Tsukuba, Japan, (7)International Institute of Tropical Agriculture, Ibadan, Nigeria, (8)International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, (9)University of Frankfurt am Main, Frankfurt, Frankfurt am Main, Germany

The availability of reference genomes is facilitating the application of genomics-assisted breeding to accelerate the genetic improvement of crops. However, the majority of crops with significant economic importance are yet to benefit from these modern tools. White Guinea yam (*D. rotundata*) is a major source of food and income for millions of people in sub-Saharan Africa. It also plays vital roles in the socio-cultural life of several societies particularly in West and Central Africa. To enhance the improvement *D. rotundata* by fully exploiting modern genetics and genomics tools, generating a reliable reference sequence is a prerequisite. To this end, an international collaborative research project was initiated by the Japan International Research Center for Agricultural Sciences (JIRCAS), Iwate Biotechnology Research Center (IBRC), and International Institute of Tropical Agriculture (IITA) to sequence the genome of *D. rotundata*. Accordingly, the genome of *D. rotundata* has been sequenced and assembled, and about 75% of the assembly was anchored and oriented using restriction-site associated DNA (RAD)-based genetic map consisting of 21 linkage groups. Significance of availability of the reference genome for developing genetic and genomic resources

for accelerating yam-breeding programs will be discussed. We strongly believe that this represents an invaluable resource to broaden our knowledge of the genetics of *D. rotundata* and provide a platform for implementing genomic-assisted breeding in this important but under-researched crop.

W900: Sweet Potato and Yam Genomics

Advanced Omic Technologies for Genetic Enhancement of Yam (*Dioscorea* spp.)

Ranjana Bhattacharjee[1], Leena Tripathi[2], Antonio Lopez-Montes[1], Michael Abberton[1], P. Lava Kumar[1], David Dekoyer[1] and Robert Asiedu[1], (1)International Institute of Tropical Agriculture, Ibadan, Nigeria, (2)International Institute of Tropical Agriculture (IITA), Nairobi, Kenya

Yam (*Dioscorea* spp.) is a multi-species clonally propagated edible tuber in the tropical and sub-tropical regions of Africa, Asia, the Pacific and the Caribbean islands. The majoity of yam productioncomes from West Africa, which accounts for about 93% of the world's production of 63 million tons from 5.3 million hectares of area harvested. The crop is a dominant source of food and income for more than 60 million people in West Africa, contributing about 300 dietary calories per person each day. Most of the farmers grow traditional landraces of yams, several of which lack resistance to pests and diseases. Of these, nematodes (*Scutellonema bradys*, *Meladoigyne* spp),yam mosaic virus (YMV), and anthracnose (*Colletotrichum gloeosporioides*) are most devastating, reducing yield by 30-90%. The yam growing countries also face the challenges of declining soil fertility as well as increased exposure to irregular rainfall and long dry spells during the growing season, which contribute to lower tuber yields.

Dioscorea rotundata (white Guinea yam) and*D. alata* (water or greater yam) are the most important among the 10 cultivated *Dioscorea* spp. Genetic enhancement of these species is constrained by a number of challenges imposed by clonal nature of the crop, inadequate knowledge on genetic diversity, long breeding cycle, low propagation ratio, and limited tools and technologies to aid breeding and selection. An integrated approach, combining advanced technologies with conventional breeding, is essential to improve efficiency and accuracy in breeding for disease resistance, food quality, and other traits. The International Institute of TropicalAgriculture (IITA) has been leading efforts with different national and international partners for more than a decade to develop and apply advanced omics techniques for yam improvement. Next-generation sequencing techniques such as genotyping-by-sequencing (GBS) and whole genome genome sequencing and re-sequencing along with metabolomics are being utilized to understand the genetic structure and relationships within and betweeb different Dioscorea species. These will facilitate genomic studies such as linkage mapping, gene/QTL identification and association mapping for marjer-assisted selection. Additionally, researchers at IITA have established a highly efficient and simple Agrobacterium-mediated transformation system for *D. rotundata*. This protocol opens up an avenue for future genetic improvement of *D. rotundata* using candidate genes for diseases and pests resistance to attain sustainable production and also provide opportunities for functional genomic studies.

W901: Sweet Potato and Yam Genomics

Progress on the Genome Characterisation of *D. alata*

Benjamen H White[1], Ranjana Bhattacharjee[2], Walter Verweij[3], Antonio Lopez-Montes[2], P. Lava Kumar[2], Sophien Kamoun[4] and **Manuel Corpas**[1], (1)The Genome Analysis Centre, Norwich, United Kingdom, (2)International Institute of Tropical Agriculture, Ibadan, Nigeria, (3)The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland, (4)The Sainsbury Laboratory, Norwich, United Kingdom

Yam is a staple crop of great agricultural, cultural and economic significance to Africa, the Americas, the Caribbean, South Pacific and Asia. Further insight into the genomics of yam species, including *Dioscorea alata*, would provide a useful resource for comparative analysis with other staple and orphan crop species around the globe. We have generated the reference genome sequences, assemblies and annotation for *D. alata* with the aim of carrying out comparative genomic analysis for yam and other related crop species. The *de novo* genome assembly and the newly created annotations will be made available through the TGAC’s resource page. All data is freely available and submitted to the relevant databases. Insight gained into the genomic content of *D alata* will open up avenues for perusing transgenic approaches in other crop species and also generating arrays for wider use in referenced and unreferenced crop species.

W902: Sweet Potato and Yam Genomics

CIRAD, IRD and INRA Yam Genomic Initiatives: Unlocking Genetic Diversity and Accelerating Yam Breeding

Hana Chaïr[1], **Nora Scarcelli**[2], Gemma Arnau[1], Claudie Pavis[3], Roland Akakpo[2], Alexandre Dansi[4], Dalila Petro[3], Karine Alix[5], Vincent Lebot[6] and Yves P. Vigouroux[2], (1)CIRAD, Montpellier, France, (2)IRD, Montpellier, France, (3)INRA Antilles-Guyane, Petit-Bourg, France, (4)University of Parakou, Parakou, Benin, (5)AgroParisTech, Gif-sur-Yvette, France, (6)CIRAD, Port-Vila, Vanuatu

The fast development of new sequencing technologies allows unlocking diversity of tropical crops. The French institutes CIRAD, IRD and INRA have worked together the last 7 years to improve the yam's genomics toolkit. Using RNASeq from leaves and flower tissues, we have assembled the transcriptomes of three yam cultivated species: *Dioscorea rotundata*, *D. alata* and *D. trifida*. Genotyping by sequencing approaches are currently used to develop a genetic map from four half-sibling families of *D. alata*. Such tools will allow refining genome assembly and identifying QTL related to anthracnose resistance. The study of genomics diversity of *D. alata* using sampling covering the four continents Asia, Pacific, Africa and America is currently underway. Using similar approaches, we obtained 600K SNPs in *D. rotundata*, *D. abyssinica* and *D. praehensilis*. We will present result of this study documenting both the relationship and diversity in these three African species and the detection of genes under selection, genes that may be of importance for yam improvement. We further investigated the important role of wild to cultivated gene flow between *D. nummularia*/*D. alata* and *D. abyssinica*/*D. praehensilis*/*D. rotundata* could play in yam improvement.

W903: Swine

Global CpG Methylation Patterns of Pigs Shows High Similarity to Other Mammals Including Humans

Min-Kyeung Choi and Chankyu Park, Department of Animal Biotechnology, Konkuk University, Seoul, South Korea

DNA methylation is a major component of the epigenetic regulation of gene expression in the mammalian genome. Although a few genome-wide DNA methylation profiling studies have been reported in pigs, the available data is still quite limited, especially for tissue diversity and detailed analysis. In order to estimate the efficiency of reduced representation bisulfite sequencing (RRBS) for the porcine genome, we carried out *in silico* analysis of the pig reference genome and compared the methylation associated genome parameters to those of humans, mice and zebra fish in which the genome level RRBS analysis results are available. Subsequently, we experimentally generated the DNA methylation profiles of the neocortex, olfactory epithelium, spleen, liver and muscle tissues and a pulmonary alveolar macrophage (PAM; 3D4/2) cell line using RRBS. On average, 3.92 Gb of clean reads were analyzed from each of the six samples. The results showed that 42.82% to 45.71% and 4.85% to 5.07% of detected cytosine and guanine dinucleotides (CpG) were located in CpG islands (CGI) and 2 kb upstream of transcription start sites (TSS), respectively. We observed a low rate (an average of 1.67%) of non-CpG methylation in the six samples except for the neocortex (2.3%). The observed global CpG methylation patterns of pigs indicated high similarity to other mammals such as humans and mice. The general characteristics of the methylation pattern of the pig genome by RRBS analysis in this study were consistent to the results of another recent pig genome RRBS study. In addition, we analyzed the correlation between the levels of DNA methylation and gene expression among neocortex, liver, muscle, spleen and a macrophage cell line. Observed patterns of differentially methylated cytosine (DMC) were compared among different tissues. These results should serve as a basic framework to establish the complete porcine methylome map and to understand the effect of DNA methylation on gene expression and phenotypic differences.

W904: Swine

Inferring Causal Gene-Phenotype Networks Underlying Complex Traits using Multi-Omics Data

Francisco Peñagaricano, University of Florida, Gainesville, FL

Association studies have been successful in identifying genomic regions associated with phenotypic traits in livestock species. However, the identification of the individual genes responsible for the phenotypic variation remains challenging. Additionally, these analyses do not provide in general any information about the molecular pathways underlying the phenotype under study. One way to unravel these molecular mechanisms is to expand the traits under analysis; one of such traits is gene expression. Indeed, the integration of phenotypic data with genotypic information and transcriptional profiling has the potential to uncover the genetic control of gene activity and phenotypic variation, as well as shed light on the manner and extent of connectedness among these variables. These connections have been explored mainly in terms of associations, i.e. connections among variables without causal interpretation. Knowledge regarding causal relationships among genes and phenotypes can be used to predict the behavior of complex systems, as well as to optimize management practices and selection strategies. Here, we describe a multistep procedure for inferring causal molecular networks underlying complex traits using multi-omics data. We initially explore marginal associations between genotypes and phenotypic and expression traits through the use of whole-genome scans, and then, in those regions where multiple significant hits co-localize, we attempt network reconstruction using causal structural learning algorithms. As a proof of principle of the significance of this integrative approach, we show the construction of causal molecular networks underlying carcass fat deposition and loin muscle weight integrating multi-omics data obtained from an F2 Duroc x Pietrain resource population.

W905: Swine

Analysis of the Boar Sperm Transcriptome Reveals RNAs Related to Spermatogenesis and Highlights Transcripts Related to Sperm's Loss of Viability

Marta Gòdia[1], Fabiana Mayer[1], Julieta Nafissi[1], Anna Esteve-Codina[2], Sam Balasch[3], Joan E. Rodríguez-Gil[4], Alex Clop[1] and Armand Sanchez[1], (1)Center for Research in Agricultural Genomics, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain, (2)Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain, (3)Grup GEPORK, Masies de Roda, Spain, (4)Unit of Animal Reproduction, Department of Animal Medicine and Surgery, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain

Insemination boars are selected based on their genetic merit for production traits. As a consequence, a proportion of these boars end up being rejected due to poor semen quality. Thus, an early prediction test of semen quality would help selecting better boars thereby improving the efficiency of insemination centers. Mature sperm contains a small population of highly degraded RNAs which might be related to semen quality and fertility. The objective of our study was to profile the boar's sperm transcriptome and identify the transcripts that are linked to the loss of spermatozoa viability (LSV) after incubation at 37ºC. Ejaculated sperm from 6 pigs with either good (n=3) or bad (n=3) LSV was subjected to RNA-seq. We then applied two computational workflows: (1) GEM+edgeR and (2) TopHat+Cufflinks and selected the 10,235 transcripts that were identified by both workflows. The 10% most highly expressed genes accounted for 98.8% of the transcript abundance. These were mostly related to spermatogenesis, ribosome assembly and translation. Differential expression analysis comparing the good and bad LSV samples highlighted 96 under- and 10 over-expressed genes (FDR < 0.05) in the good LSV spermatozoa. Some of these genes are related to male infertility and abnormal spermatogenesis according to the Mouse Genome Informatics database. In keeping with studies on other mammals, our results suggests a high specialization of the boar's sperm transcriptome in which most abundant RNAs are related to spermatogenesis and fertility and relates some transcripts to LSV.

W906: Swine

Quality Assessment of the Current Pig Genome for Variant Discovery; and High Quality Re-assembly Using 3rd Generation Sequencing

Amanda Warr[1], Christelle Robert[1], David A. Hume[2], Alan L. Archibald[2], Joseph (Nader) Deeb[3] and Mick Watson[1], (1)The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom, (2)The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, United Kingdom, (3)Genus plc, Hendersonville, TN

Many applications of high throughput sequencing rely on the availability of an accurate reference genome. Variant calling often produces large data sets that cannot be realistically validated and which may contain large numbers of false-positives. Errors in the reference assembly increase the number of false-positives. While resources are available to aid in the filtering of variants from human data, for other species these do not yet exist and strict filtering techniques must be employed which are more likely to exclude true-positives. This work assesses the accuracy of the pig

reference genome (Sscrofa10.2) using whole genome sequencing reads from the Duroc sow whose genome the assembly was based on. Indicators of structural variation including high regional coverage, unexpected insert sizes, improper pairing and homozygous variants were used to identify low quality (LQ) regions of the assembly. Low coverage (LC) regions were also identified and analyzed separately. The LQ regions covered 13.85% of the genome, the LC regions covered 26.6% of the genome and combined (LQLC) they covered 33.07% of the genome. Over half of dbSNP variants were located in the LQLC regions. Of copy number variable regions (CNVRs) identified in a previous study, 86.3% overlapped with the LQLC regions. Researchers using WGS data should be aware that the current pig reference genome does not give an accurate representation of the copy number of alleles in the original Duroc sow's genome. We will also present data from a new assembly of the pig genome assembled de novo from 65X coverage of TJ Tabasco. We will highlight improvements made compared to the existing assembly and outline a path to the next generation of the pig genome.

W907: Swine

What Factors Jeopardize Whole Transcriptome Analysis using Next Generation Sequencing Approaches?

Zhihua Jiang[1], Xiang Zhou[1], Jennifer J. Michal[1], Rui Li[1] and Richard M. Harland[2], (1)Washington State University, Pullman, WA, (2)University of California Berkeley, Berkeley, CA

Whole transcriptome analysis begins with preparation of next generation sequencing libraries from high quality total RNA or purified mRNA samples, followed by sequencing using the state-of-art high throughput sequencers and data processing and analysis. During development of our whole transcriptome target sequencing (WTTS) method to capture 3' ends of transcripts, we observed several factors that jeopardized whole transcriptome analysis. Inappropriate primer design resulted in both recessive and dominant "amplification detours" that produced noisy reads and biased data. In addition, excessive PCR cycles and high concentrations of primers resulted in over-amplification of next generation sequencing libraries and reduced transcriptome coverage. Incomplete genome sequencing and partial gene annotation results in missing data, particularly when a 3' end sequencing approach is employed. Overlapping genes can transfer reads from one expressed gene to another non-expressed gene or cause non-strand-restricted reads (such as paired-end reads) to be unmapped. RNA-seq tends to enlarge whole transcriptome up-bottom boundaries, producing a much longer list of differentially expressed genes in comparison to the WTTS method. Furthermore, RNA-seq reads cannot be used to detect polyA sites or potential isoform switches in tissues at different time points. When an Illumina analyzer is used for sequencing, libraries with low diversity must be avoided. We will demonstrate cases associated with these challenges at the Workshop.

W908: Swine

Probing Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Infection Control Mechanisms using Differential Gene Expression

Joan K. Lunney[1], Igseo Choi[2], Hua Bao[3], Arun Kommadath[3], Le Luo Guan[3], Graham S Plastow[4], Raymond R. R. Rowland[5], Sam M. Abrams[1], James M. Reecy[6], Eric Fritz-Waters[7], Christopher K. Tuggle[6], Jack C.M. Dekkers[6] and Paul Stothard[3], (1)APDL, BARC, ARS, USDA, Beltsville, MD, (2)APDL BARC ARS USDA, Beltsville, MD, (3)Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada, (4)University of Alberta, Edmonton, AB, Canada, (5)Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, KS, (6)Iowa State University, Ames, IA, (7)Department of Animal Science, Iowa State University, Ames, IA

Porcine reproductive and respiratory syndrome (PRRS) is economically the most important disease of pigs with annual U.S. losses of $664M. Understanding transcriptional control of anti-viral responses should reveal mechanisms to help control PRRS. The PRRS Host Genetics Consortium (PHGC) was established to combine efforts of scientists from university, government and commercial pig genetics and animal health companies to assess the role of genetics in determining pig resistance/susceptibility to PRRS virus (PRRSV) infection, pathology and growth effects. The PHGC used a nursery pig PRRSV infection model with deep sampling for phenotypic analyses, extensive genotyping (60K SNPchip) and a shared database http://www.animalgenome.org/lunney/. To address disease resistance mechanisms we probed the blood transcriptome of PHGC pigs using RNAseq. Based on data from 14 trials using ~200 PRRSV-infected pigs each we identified a genomic region on SSC4, and a putative candidate gene GBP5, which had significant impact on variation in viral load and growth responses following challenge with each of 2 different PRRSV isolates. Using RNAseq analyses we have identified genes that are differentially expressed in PRRS resistant versus susceptible pigs. Our current efforts are aimed at probing for alternate control and regulatory networks by comparing pigs with differential PRRSV responses as well as assessing interactions between viral and host genes. This data has already enabled us to identify expression QTL (eQTL) and should reveal new pathways that may be used for design of novel vaccines and biotherapeutics. Support: US National Pork Board, USDA ARS and NIFA, Genome Canada, Genome Alberta, pig breeding companies.

W909: Swine

Do you really know where this SNP goes?

Gary A. Rohrer[1], Dan Nonneman[1], Steven G. Schroeder[2], Jason Chin[3], Sergey Koren[4], Adam Phillippy[4], Richard Green[5], Nicholas Putnam[6] and Timothy P.L. Smith[1], (1)USDA, ARS, USMARC, Clay Center, NE, (2)Animal Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, (3)Pacific Biosciences, Menlo Park, CA, (4)National Human Genome Research Institute, Bethesda, MD, (5)University of California, Santa Cruz, Santa Cruz, CA, (6)Dovetail Genomics, Santa Cruz, CA

The release of build 10.2 of the swine genome was a marked improvement over previous builds and has proven extremely useful. However, as most know, there are regions of the genome that this particular build does not accurately represent. For instance, nearly 25% of the 62,162 SNP on the Illumina Porcine SNP60v1 BeadChip do not map to a unique position in build 10.2. To resolve these issues of the swine genome, we utilized advances in long-read sequencing combined with new technologies for scaffolding the resulting contigs to create a de novo build of the genome for a male pig from the USMARC swine population. A total of 85% of the SNP sequences uniquely mapped to the new genome assembly. More importantly, when comparing SNP genomic position between both builds, several inconsistencies arose. Eighty inconsistencies involving assignments to different chromosomes or large discrepancies in location (> 20 Mb) were investigated. These inconsistencies were evaluated by comparing linkage disequilibrium values between subsets of markers involved in these inconsistencies with flanking markers of both predicted regions. Linkage disequilibrium values were computed from genotypes collected on over 3,000 pigs in the USMARC swine

commercial population. A total of 223 SNP markers were studied and 22% of the inconsistencies could not be resolved, 2% agreed with build 10.2 and 76% of the inconsistencies indicated that the new build was correct.
USDA is an equal opportunity provider and employer.

W910: Swine

Molecular Cloning of WAS Protein Family, Member 1 (WAVE1) in Porcine Oocytes

Kiho Lee, Luke Dillard and Junghyun Ryu, Virginia Tech, Blacksburg, VA

Oocytes have ability to remodel epigenetic marks of somatic cells therefore reprogram the somatic cells into embryonic state. Unique reprogramming factors which reside in oocytes are thought to be contributing to this reprogramming process. However, specific oocytes factors are not well-characterized. The *WAVE1* gene, derived from Xenopus eggs, is shown to reprogram murine somatic cells into pluripotent state. The gene is specifically related to transcriptional activation of embryonic genes by affecting RNA polymerase II activity. When amino acid sequence of WAVE1 was compared among various species, it was highly conserved although little similarity was found at the nucleotide level. Because the gene is highly conserved, we hypothesized that the gene has a similar role in pigs as a reprogramming factor. To clone the full coding sequence of porcine *WAVE1*, we identified a porcine homolog of *WAVE1* from GenBank. RT-PCR results using primers designed based on the sequence demonstrate that the *WAVE1* is expressed in porcine oocytes. To clone the entire coding sequence of porcine *WAVE1*, predicted sequence of porcine *WAVE1* and ESTs extending sequences of the predicted *WAVE1* were identified from GenBank. Primers designed based on the sequence information could amplify the entire coding sequence of porcine *WAVE1* from cDNA, derived from oocytes. Sequencing results indicate that the porcine WAVE1 has high identity with human and Xenopus at the amino acid level; 99% and 82% respectively. The coding sequence of porcine WAVE1 will be overexpressed in somatic cells to examine its ability to reprogram somatic cells in pigs.

W911: Swine

Identification of Runs of Homozygosity in Large White and Landrace Sows Associated with Decreased Number of Piglets Born

Christian Maltecca[1], Jeremy Howard[1], Kent Gray[2], Francesco Tiezzi[3] and YiJIan Huang[2], (1)NCSU, Raleigh, NC, (2)Smithfield Premium Genetics, Rose Hill, NC, (3)NCSU, raleigh, NC

Utilizing dense genotypes in swine genetic evaluations has become a routine practice. A supplemental use of the genotype information can be to identify regions of the genome that give rise to inbreeding depression. The objective of the study is to identify regions when in a run of homozygosity (ROH) impact the number of piglets born alive in a Landrace (LR) and Large White (LW) pig population. Genotyped dams (n = 5000 LR; n = 5000 LW) with number born alive (NBA) phenotypes were utilized to construct yield deviations. A ROH statistic (ROH5Mb), which is whether a SNP was in a ROH of 5 Mb, was calculated across the genome. Utilizing a Bayesian Ridge Regression approach yield deviations weighted by the information content were regressed on the additive and ROH5Mb effect of a SNP along with the polygenic effect of the dam. Regions that had a large variance based on 1-Mb sliding window genomic estimated breeding values were investigated further in order to determine the phenotypic effect. For each ROH investigated, the actual phenotype was used as phenotypes in an analysis along with fixed effects of contemporary group, parity, SNP contained within the ROH, ROH and random effects of dam and the permanent environmental effect of the dam. Regions on SSC1, SSC6 and SSC13 for LR and SSC4, SSC13 and SSC 17 for LW caused a significant reduction (ranged from -0.64 to -0.34) in the NBA. These regions can be utilized in mating programs to reduce the impact of inbreeding depression.

W912: Swine

Genomic Differences Between Pre-weaning Survival and Mortality of Piglets Following PEDv Outbreaks

Francesca Bertolini[1], John Harding[2], Benny Mote[3], Graham S Plastow[4] and Max F. Rothschild[1], (1)Iowa State University, Ames, IA, (2)University of Saskatchewan, Saskatoon, SK, Canada, (3)University of Nebraska-Lincoln, Lincoln, NE, (4)University of Alberta, Edmonton, AB, Canada

PED (Porcine Epidemic Diarrhea) is a disease caused by PED virus, which belongs to the Coronavirus family and affects pigs of all ages, but can cause 100% mortality in suckling piglets in the acute phase of the disease. This disease has been relatively common in Asia and Europe since the 1970s, while the USA experienced a widespread epidemic of PED beginning in 2013. In this research, 156 dead and 106 surviving pigs from five PED outbreak farms located in USA, Canada or Germany were genotyped with the 80K SNP chip. Sampling occurred within 3 weeks of the initial onset of PED, prior to the development of maternal immunity. Surviving piglets experienced PED diarrhea but did not succumb to the disease during lactation. After filtering the marker data 60,969 SNPs were used in an Fst analysis comparing dead and surviving pigs using non-overlapping 1Mb windows (using the average Fst of all the SNPs in each window). Results were then normalized to find regions that diverged more than others within the comparison. Significant regions associated with survival were found on chromosomes 2, 4 and 15 and contained 162 genes. Analysis with EnriChr showed that several genes were under or over expressed in several coronavirus infections in humans. Analysis with GOrilla revealed enrichment for genes that belong to the Golgi apparatus and membrane, together with genes involved in ion transporter, metabolism, and ATPase activity.

W913: Swine

UIUC Research Update

Jonathan E. Beever, Department of Animal Sciences, University of Illinois at Urbana-Champaign, Urbana, IL

W914: Swine

Development of Resources and Tools for Mapping Genetic Sources of Phenotypic Variation

Daniel C. Ciobanu[1], Emily Tosky[1], Sean Olson[1], Melanie Trenhaile[1], Clay A Lents[2], Timothy P.L. Smith[2], Dan Nonneman[2], Gary A. Rohrer[2], Jason Chin[3], Phillip S. Miller[1], Thomas Burkey[1], Matt L. Spangler[1], Jean-Jack Riethoven[1], Graham S Plastow[4] and

Stephen D. Kachman[1], (1)University of Nebraska - Lincoln, Lincoln, NE, (2)USDA, ARS, USMARC, Clay Center, NE, (3)Pacific Biosciences, Menlo Park, CA, (4)University of Alberta, Edmonton, AB, Canada
Commercial and experimental genetic resources were established and investigated for a range of reproductive and disease susceptibility phenotypes. These efforts were complemented with RNA and whole genome sequencing and new assemblies of the swine genome. Data generation efforts were accompanied by the development of novel statistical approaches, integrating haplotype analysis and interval mapping in a Bayesian framework to explore genetic variation of multiple traits.
Specifically, sources of variation in expression of puberty and fertility were investigated by RNAseq of hypothalamic arcuate nucleus, various GWAS approaches, and genome sequencing of sires with early and late puberty daughters. Various subsets of SNPs from major 1-Mb windows that explained the largest proportions of variation in training populations were evaluated to assess the potential of reduced sets of SNP to explain phenotypic variation for age at puberty. Relative to disease, a major QTL for PCV2 viremia previously reported was fine mapped on a novel 29 Mb scaffold based on long sequencing reads while gene prediction and annotation provided potential candidates responsible for the observed variation. In addition, new PCV2 challenges targeting pigs of different QTL genotypes in experimentally infected and vaccinated animals confirmed previous results and provided evidence of the host genetic role in viral replication.

W915: Swine

Post-Weaning Blood Transcriptomic Differences Between Yorkshire Pigs Divergently Selected for Residual Feed Intake

Haibo Liu[1], Nguyen T. Yet[2], Dan Nettleton[2], Jack C.M. Dekkers[3] and **Christopher K. Tuggle**[3], (1)Bioinformatics and Computational Biology Program, Department of Animal Science, Iowa State University, Ames, IA, (2)Department of Statistics, Iowa State University, Ames, IA, (3)Department of Animal Science, Iowa State University, Ames, IA
Improving feed efficiency (FE) of pigs by genetic selection is of economic and environmental significance. One measure of feed efficiency is residual feed intake (RFI). Currently, the molecular mechanisms underlying RFI are largely unknown. In this study, we explored whether differences exist in the global gene expression profiles of peripheral blood of 35-42 day-old pigs with extremely low (more efficient) and high RFI from two lines that were divergently selected for RFI, to use such information to explore the molecular basis of RFI differences, and to initiate development of predictive biomarkers for RFI. We identified 454 differentially expressed genes (DEGs) ($q \leq 0.05$) between the low (n = 15) and high (n = 16) RFI animals by using RNAseq. We validated 24 of 37 selected DEGs by RT-qPCR. The 1972 DEGs ($q \leq 0.15$) significantly overlapped with genes associated with several diseases, including hyperphagia, eating disorders and mitochondrial diseases (q < 1E-05). Gene Ontology term overrepresentation analysis suggested differences in mitochondrial and proteasomal activities, small molecule biosynthetic process, and signal transduction between the two RFI lines and provided new insights into the molecular basis of RFI in pigs. A weighted gene co-expression network analysis identified four co-expression modules differentially expressed between the low and high RFI groups, and there was significant overlap between cluster membership and DEGs. These DEGs and representative genes from co-expression modules that were associated with RFI phenotype provided a preliminary list for developing predictive biomarkers for RFI in pigs.
Supported by USDA- NIFA-AFRI-2011-68004-30336.

W916: Swine

A Preliminary Analysis of *Longissimus* Muscle microRNA Expression Profiles in the Michigan State University F2 Duroc X Pietrain Pig Resource Population

Catherine W. Ernst[1], Kaitlyn Perry[1], Juan P. Steibel[1], Deborah Velez-Irizarry[1], Scott A. Funkhouser[2], Sebastian Casiro[1], Nancy E. Raney[1] and Ronald O. Bates[1], (1)Department of Animal Science, Michigan State University, East Lansing, MI, (2)Genetics Program, Michigan State University, East Lansing, MI
MicroRNAs (miRNAs) are a class of small, single-stranded RNAs shown to post-transcriptionally regulate gene expression through complementary binding with target mRNAs. While miRNAs have been shown to regulate gene expression during fetal and early postnatal pig skeletal muscle development, miRNA regulation in market-age pigs has been less explored. The overall objectives of this study are to profile miRNAs in *Longissimus dorsi* (LD) muscle of F2 pigs from the MSU Duroc x Pietrain resource population for integration into a genome-wide eQTL analysis, and to elucidate gene networks controlling economically important phenotypes. Total RNA extracted from 147 LD samples was sequenced on the Illumina HiSeq 2500 platform. Raw sequence reads were trimmed, quality-filtered and PCR duplicates were removed. After filtering for low read counts, 241 normalized annotated mature miRNA expression profiles were included in a preliminary GBLUP analysis to identify miRNA exhibiting heritable expression utilizing the gwaR package. These miRNA with heritable expression are strong candidates for exhibiting association with genome-wide SNP markers. Average heritability of the 241 miRNAs was 0.16, whereas heritability of the 69 significant miRNAs was 0.34 ($q < 0.05$). Additionally, miRDeep2 was utilized to predict potentially pig-novel miRNA precursors revealing 235 putative precursors at an estimated false positive rate of 11%. Of these precursors, 89 have a homologous approximately 6-nt seed sequence with annotated human mature miRNA, suggesting their potential as putative pig-novel miRNAs. Further investigation is required to characterize these miRNAs, in order to identify candidates for inclusion in future analyses. Supported by USDA-NIFA-AFRI Grant 2014-67015-21619.

W917: Synthetic Biology

Field performance of sugarcane with TALEN mediated targeted mutagenesis of COMT

Baskaran Kannan, Je Hyeong Jung and Fredy Altpeter, University of Florida - IFAS, Gainesville, FL
Transcription activator-like effector nuclease (TALEN) is a recently developed tool enabling precise genome modifications, such as targeted mutagenesis, gene replacement, or insertion. Sugarcane is a prime feedstock for bioethanol production, and utilizing both sucrose and cell wall bound sugars for fermentation will enhance the biofuel yield. We recently demonstrated that RNAi mediated downregulation of lignin biosynthetic gene Caffeic acid O-methyltransferase (COMT) is a successful strategy to improve bioethanol production from lignocellulosic sugarcane biomass. In this study, COMT was targeted for the TALEN induced multi-allelic mutagenesis to modify lignin biosynthesis in sugarcane. Targeted mutations following TALEN delivery were identified by capillary electrophoresis of the COMT amplicon. Events were confirmed by sequencing of the COMT amplicon which revealed the presence of insertions and deletions at the target site. Data comparing the total lignin content in the stem biomass of COMT mutant and RNAi suppressed sugarcane events and their field performance will be presented.

W918: Synthetic Biology

A Systems Biology Approach to Understand Terpene Biosynthesis in *Eucalyptus* for Advance Biofuel Engineering

Ritesh Mewalal, Timothy J. Tschaplinski and Gerald A. Tuskan, Oak Ridge National Laboratory, Oak Ridge, TN

Terpenes are naturally occurring, chemo-diverse set of compounds commonly synthesized within foliage in plants. These compounds are hydrocarbons categorized by the number of recurring isoprene units (C5H8), e.g., mono- (C10), sesqui- (C15), di- (C20) and triterpenes (C30). The large range of reported chemical structures is paralleled with a number of biological roles in plants and prospective industrial utility. Specific terpenes have the potential to meet the physicochemical requirements of current fuel technology including, low temperature viscosity, high energy density and volumetric net heat of combustion. β-pinene, α-pinene, linalool, farnesene, camphene, limonene, 1,4-cineole, 1,8-cineole, sabinene, bisabolene, and myrcene have been identified as promising advanced biofuel candidates. These terpenes are synthetized with the oil glands of foliage tissue of *Eucalyptus* and vary in their presence and concentration among different species. The aim of this study is to use system biology, integrating (and contrasting) metabolome, transcriptome and proteome data from an ontological developmental foliar growth series (and oil glands) to understand the genetic regulation of terpene biosynthesis in *Eucalyptus.* We will be presenting the preliminary data of species and genotype selection based on metabolite profiling, gland size and number of glands per leaf. In addition, we will show initial transcriptomic analysis of the oil glands along a leaf developmental series contrasting against the whole-leaf transcriptome from these stages. This study will have important implications for future biotechnology and synthetic biology by revealing promising candidate biosynthetic enzymes, transporter and regulatory proteins.

W919: Synthetic Biology

Application of CRISPR/Cas9 System for Modification of Flower Colors in Tobacco and Torenia Plants

Masahiro Nishihara, Atsumi Higuchi, Kohei Fujita, Keisuke Tasaki, Yoshimi Kurokawa, Hideyuki Takahashi and Nobuhiro Sasaki, Iwate Biotechnology Research Center, Kitakami, Iwate, Japan

Genome modification technologies such as ZFN (Zinc Finger Nuclease), TALEN (Transcription Activator-Like Effector Nucleases) and CRISPR/Cas9 technology have been successfully applied to improve various crop traits. Among them, CRISPR/Cas9 system is now conveniently available as a precise and efficient genome engineering tool. Targeted mutagenesis using CRISPR/Cas9 system is especially useful for plant breeding and gene functional analysis.

In this study, we applied the CRISPR/Cas9 system to modify flower colors in higher plants using tobacco (*Nicotiana tabacum* cv. SR1) and torenia (*Torenia fournieri*) as models. Binary vectors harboring expression cassettes of Cas9 nuclease, single guide RNA (sgRNA) targeting the flavanone 3-hydroxylase (*F3H*) and *nptII* as a selectable marker gene were constructed and used for production of transgenic tobacco and torenia plants via *Agrobacterium*-mediated transformation. Targeted mutations were analyzed in kanamycin-resistant shoots by sequencing of the *F3H* amplicons (PCR products). Transgenic tobacco and torenia plants clearly contained DNA mutations such as nucleotide substitutions, insertions and deletions at the target site, which varied depending on the transgenic lines. Flower colors of some transgenic lines turned pale or white. Interestingly, some lines displayed variegated flowers. Data describing relationship between flower-color phenotypes and mutations in these transgenic plants will be presented.

W920: Synthetic Biology

RNA Editing in the Bovine Transcriptome

Deepak R. Unni[1], Darren E. Hagen[1] and Christine G. Elsik[1,2], (1)Division of Animal Sciences, University of Missouri, Columbia, MO, (2)MU Informatics Institute, University of Missouri, Columbia, MO

RNA editing is a post-transcriptional process that results in altered transcript sequences due to the modification of nucleotides. Of the known types of RNA editing, the most common is the deamination of adenosine to inosine (A-to-I), catalyzed by adenosine deaminase acting on RNA (ADAR). These nucleotide substitutions occur in a site-specific and tissue dependent manner and can lead to increased protein sequence diversity and altered UTR binding sites. A-to-I editing can be observed as an adenosine to guanine (A-to-G) substitution in RNA with respect to the DNA. We searched for putative RNA editing sites in *Bos taurus* using high throughput RNAseq data of 48 tissues harvested from L1 Dominette 01449, the bovine reference genome individual. RNAseq reads (Illumina,100 bp single-end) were aligned to the bovine UMD3.1 genome assembly using BWA and Hisat. Reads with high quality bases were used to identify variants; then RNAseq variants were filtered for single nucleotide polymorphisms (SNP) using dbSNP and SNPs called from the alignment of ~60x Dominette genomic sequencing data. A variant was predicted if it was represented in at least ten reads and greater than 10 percent of the reads supported the alternate allele. Our analysis revealed hundreds of A-to-I RNA editing sites in the bovine transcriptome. We predicted the effects of edits within protein coding regions, and whether editing in UTR impact miRNA binding sites.

W921: Systems Biology and Ontologies

Planteome: A resource for Common Reference Ontologies and Applications for Plant Biology

Pankaj Jaiswal[1], Laurel Cooper[1], Justin L. Elser[1], Austin Meier[2], Marie-Angélique Laporte[3], Christopher Mungall[4], Barry Smith[5], Ethan KS. Johnson[2], Megan Seymour[2], Justin Preece[6], Xu Xu[2], Ryan S. Kitchen[2], Botong Qu[2], Eugene Zhang[2], Elizabeth Arnaud[3], Seth Carbon[7], Sinisa Todorovic[2] and Dennis Wm. Stevenson[8], (1)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (2)Oregon State University, Corvallis, OR, (3)Bioversity International, Montpellier Cedex 5, France, (4)Berkeley Bioinformatics Open-Source Projects, Berkeley, CA, (5)University at Buffalo, Buffalo, NY, (6)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (7)Lawrence Berkeley National Laboratory, Berkeley, CA, (8)New York Botanical Garden, Bronx, NY

Around the world, a small number of plant species serve as the primary source of food for the human population, yet these crops are vulnerable to multiple stressors, such as diseases, nutrient deficiencies and unfavorable environmental conditions. Traditional breeding methods for plant improvement may be combined with next-generation methods such as automated scoring of traits and phenotypes to develop improved varieties. Linking these analyses to the growing corpus of genomics data generated by high-throughput sequencing, transcriptomics, proteomics,

phenomics and genome annotation projects requires common, interoperable, reference vocabularies (ontologies) for the description of the data. The Planteome initiative (www.planteome.org) is developing the needed suite of common reference ontologies that describe anatomy and development in Plant Ontology (PO), Plant Trait Ontology (TO), and Plant Environment Ontology (EO) and the Plant Stress Ontology (PSO) for describing the abiotic and biotic stresses. The project will also host ontologies such as Gene Ontology (GO), Chemical Entities of Biological Interest (ChEBI), Protein Ontology (PRO) and the Phenotypic Qualities Ontology (PATO) developed by collaborators. The project database and the online resources will provide researchers tools to search and browse and remote access via APIs for semantic integration in annotation tools and data repositories providing resources for plant traits, phenotypes, diseases, genomes, gene expression and genetic diversity data across a wide range of plant species. The project is supported by the National Science Foundation award IOS #1340112

W922: Systems Biology and Ontologies

NDEx: Sharing and Publishing Biological Networks for any Species

Dexter Pratt, UC San diego, La Jolla, CA

NDEx, the Network Data Exchange is an online commons where scientists can share and publish biological networks in many formats. NDEx users manage access to the networks that they store, keeping some private, sharing others with collaborators or making them publicly accessible. NDEx provides a venue in which researchers can create and distribute species-specific networks. NDEx also provides a versatile API to make networks accessible to software, from simple scripts to Cytoscape and other analytic applications. Finally, this novel infrastructure promotes the development of new applications and services such as viewers to embed 'live' networks directly in peer-reviewed publications. NDEx is an important step towards an ecosystem in which crowd curation allows networks bearing data, hypotheses, and findings to flow easily between scientists. Live, hands-on examples will be shown.

W923: Systems Biology and Ontologies

Functional Annotation at Scale: Pipelined Analysis of Phytozome Data and Results from Large Inter-Species Comparisons.

Joseph W. Carlson, David M. Goodstein, Richard D Hayes, Shengqiang Shu, Jeremy L. Phillips and Daniel S. Rokhsar, DOE Joint Genome Institute, Walnut Creek, CA

Phytozome, the Joint Genome Institute's portal for Comparative Plant Genomics, hosts ~60 assembled and annotated plant genomes, drawn from a broad range of Viridiplantae species with relevance to biofuels, climate change adaptation, and plant evolution. Approximately two-thirds of these genomes were assembled and/or annotated at the JGI, with the remainder acquired from external sources.

As a centralized integrated database, Phytozome enables large scale classification and analysis of plant genomic data using consistent, versioned and well-defined protocols. We have recently incorporated InterProScan v5 and the PlantCyc system into our automated functional analysis workflows, using the Nextflow system to handle pipelining and parallelization tasks. Nextflow allows us to bypass certain issues with deploying InterProScan on SGE-based compute clusters, and provides a uniform workflow platform for most of our computational pipelines.

Together with our systematic computation of gene orthologs between organism pairs and our RNA-seq based expression pipelines, we provide a rich set of comparative tools and data for the functional annotation of all Phytozome proteins. We illustrate the utility of the database for functional annotation through the use of gene expression, orthology, and pathway mapping to compare imputed function in both model and non-model plant systems.

W924: Systems Biology and Ontologies

QTLNetMiner - Linking Crop Traits to Genes through Data Integration and Text Mining

Keywan Hassani-Pak, Ajit Singh and Christopher Rawlings, Rothamsted Research, Harpenden, United Kingdom

Correct identification of causative genes for agronomic traits is valuable for effective marker assisted breeding and reverse genetics. The systematic evaluation of potential functional candidates is, however, time-consuming and requires the integration of many different types of information. These include a range of ontologies for phenotypes, gene function annotations and biochemical pathways and also gene expression data and comparative information from related organisms. An important source of information is gene and phenotype data mined from the scientific literature. We have developed data integration workflows for building genome-scale knowledge networks using the free and open-source Ondex data integration framework (www.ondex.org). QTLNetMiner is a modern web application to interrogate such large genome-scale knowledge networks with trait-based search terms (e.g. early flowering, disease resistance), QTL and gene lists. It allows users to identify and rank potential candidate genes by combining their own data with information from public databases. The user is guided and supported when writing the search terms through features such as real-time user feedback and ontology-based query term suggestions. The relevance of a gene to a particular search query is evaluated using information retrieval methods that take into account the amount of supporting evidence and the specificity of evidence to this gene. The supporting gene-evidence networks are visualized in the KNETViewer which is optimised for visualising and exploring data rich knowledge networks. QTLNetMiner instances for several plant and animal species are available here: https://ondex.rothamsted.ac.uk/QTLNetMiner/.

W925: Systems Biology and Ontologies

An Integrated System-Wide Maize Atlas: From Transcriptome to Proteome Networks

Justin W. Walley[1], Ryan C Sartor[2], Robert Schmitz[3], Zhouxin Shen[4], Joseph Ecker[5] and Steven Briggs[4], (1)Iowa State University, Ames, IA, (2)UCSD, La Jolla, CA, (3)University of Georgia, Athens, GA, (4)University of California, San Diego, La Jolla, CA, (5)Salk Institute for Biological Studies & Howard Hughes Medical Institute, La Jolla, CA

Integrated molecular atlases make possible systems biology approaches aimed at understanding biological phenomena. Using RNA-seq and quantitative mass spectrometry we generated an atlas comprised of 62,547 mRNAs, 17,862 non-modified proteins, and 6,227 phosphoproteins harboring 31,595 phosphorylation sites, quantified across maize development. Analysis of the atlas has revealed complex spatiotemporal patterns of gene activity. For example, there is poor correlation between protein and mRNA levels and for many of the most abundant proteins there is little to no detectable cognate mRNA. The atlas has also enabled generation and interrogation of a number of different types of regulatory

networks including mRNA and protein co-expression networks, and gene regulatory networks (GRN). Together, these studies highlight the complex interplay of transcriptional, translational and post-translational events in proteome dynamics.

W926: Systems Biology and Ontologies

Rice Stress-Response Gene-Network: An Example of Community Curation in WikiPathways for Plants

Sushma Naithani[1], Christina Partipilo[1], Bijayalaxmi Mohanty[2], Dong-Yup Lee[2] and Pankaj Jaiswal[1], (1)Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, (2)National University of Singapore, Singapore, Singapore

In recent decades, genome-scale expression studies have allowed identification of genes and gene-networks associated with both biotic and abiotic stress responses in plants. However, an interactive, curated model of a plant's stress-response network is lacking. We created rice stress-response interactome consisting of 500 rice (Oryza sativa) genes using WikiPathways based on published studies and analysis of transcriptomic data sets available in the public domain. We identified set of up- and down-regulated network hubs that are shared between rice biotic and abiotic stress-response subnetworks. The rice stress-response networks described herein, is freely available on the plants Wikipathways portal (http://www.wikipathways.org/index.php/Portal:Plants). Researchers can further curate components of this gene network, links to external resources, and add relevant information as new knowledge becomes available. Furthermore, users can edit/visualize this network model online and conduct analysis of their expression data using the expression viewer and statistical analysis features built into the desktop version of the PathVisio. We invite researchers to register free of charge at Wikipathways and help curate important plant pathways.

W927: Systems Biology and Ontologies

Systems Biology Approach to Understand the Rice Plant Metabolic and Transcriptional Regulation Under Abiotic Stress

Meiyappan Lakshmanan[1], **Bijayalaxmi Mohanty**[2] and Dong-Yup Lee[1,2], (1)Bioprocessing Technology Institute, Singapore, Singapore, (2)National University of Singapore, Singapore, Singapore

Rice plants are exposed to a wide range of abiotic stresses during their cultivation, that seriously affects the annual global rice production. Thus, a comprehensive understanding on how rice plants respond to these changing environments both at metabolic and gene regulation levels will provide a significant insight towards improved rice varieties. To do so, we employed systems biology approach, and initially developed a core mathematical model of rice which allows us to characterize cellular behavior and metabolic states under various abiotic stress conditions, such as i) photorespiratory pathway in rice leaves and identification of essential and lethal genes of the pathway , and ii) metabolically and transcriptionally regulated reactions and potential transcription factors involved in the regulation of coleoptile germination and elongation of rice seeds under anoxia. The core model was then further expanded to reconstruct a fully compartmentalized genome scale metabolic model. Subsequently, transcriptomics and metabolomics data were systematically integrated with the model to identify the potential transcription factors, i) in leaf, root and panicle tissues at three different developmental stages in response to drought stress , and ii) the control of light-mediated signaling mechanisms. The information derived from the current in silico analysis in conjunction with multi-omics profiling can potentially guide for developing new breeding and/or engineering targets as crop productivity strategies.

W928: Systems Biology and Ontologies

AgroPortal : A Proposition for Ontology-Based Services in the Agronomic Domain

Clement Jonquet, Laboratory of Informatics, Robotics, and Microelectronics of Montpellier (LIRMM), Montpellier, France, Esther Dzalé-Yeumo, French National Institute for Agricultural Research (INRA), Versailles, France, Elizabeth Arnaud, Bioversity International, Montpellier, France and Pierre Larmande, IRD, UMR DIADE, Institut de Biologie Computationnelle, Montpellier, France

Our project is to develop and support a reference ontology repository for the agronomic domain. By reusing the NCBO BioPortal technology, we have already designed and implemented a prototype ontology repository for plants and a few crops. We plan to turn that prototype into a real service to the community. The AgroPortal project aims at reusing the scientific outcomes and experience of the biomedical domain in the context of plant, agronomic and environment sciences. We will offer an ontology portal which features ontology hosting, search, versioning, visualization, comment, but we will also offer services for semantically annotating data with the ontologies, as well as storing and exploiting ontology alignments and data annotations. All of these within a fully semantic web compliant infrastructure. The main objective of this project is to enable straightforward use of agronomic related ontologies, avoiding data managers and researchers the burden to deal with complex knowledge engineering issues to annotate the research data. The AgroPortal project will specifically pay attention to respect the requirements of the agronomic community and the specificities of the crop domain. We will first focus on the outputs of a few existing driving agronomic use cases related to rice and wheat, with the goal of generalizing to other Crop Ontology related use cases. AgroPortal will offer a robust and stable platform that we anticipate will be highly valued by the community.

W929: Systems Genomics

Transcriptome and Proteome Derived Networks Reveal Distinct and Complementary Gene Relationships

Steven Briggs, University of California, San Diego, La Jolla, CA

W930: Systems Genomics

Systems Analysis of 1,090 *GSTR* Genes Reveals That Epistasis Plays Important Roles in Cotton Fiber Strength

Yun-Hua Liu, Meiping Zhang, Yang Zhang, C. Wayne Smith, Steve Hague and Hong-Bin Zhang, Texas A&M University, College Station, TX

Fiber bundle strength is one of the paramount quality traits in the world’s textile and yarn industry, and stronger fiber tenacity promises better quality yarn and thus final products. Nevertheless, little is known about the molecular mechanisms underlying fiber strength in cotton. Here we report 1,090 genes controlling fiber strength, named as *GSTR* for *Gossypium* fiber strength, cloned using a novel high-throughput gene cloning system developed in our laboratory. All of the genes have been validated and their effects on fiber bundle strength have been estimated by

regulating their activities in multiple cotton lines. Each of the genes has been shown to have an effect varying from 3.64 – 14.16% on fiber strength. However, 126 (11.56%) of them led to stronger fibers, whereas 964 (88.44%) led to weaker fibers when they were turned on or up-regulated. The *GSTR* genes encode proteins and enzymes involved in a variety of biochemical reactions, biological processes and metabolic pathways. We show that 1,090 *GSTR* genes function collaboratively and constitute an integrated co-expression network with 250,202 gene x gene or epistatic interaction edges. Although every *GSTR* gene was maintained in the network, the interactions of the *GSTR* genes varied dramatically across different cotton lines and were significantly associated with fiber strength. These results suggest that fiber strength is not only determined by genes, but also largely dependent on the magnitude, composition and number of their epistatic interactions, thus providing a first insight into the molecular mechanisms of fiber strength and enabling toolkits for enhanced cotton breeding.

W931: Systems Genomics

Large-scale Gene Association Network Inference and Functional Module Discovery in Plants

Patrick Xuechun Zhao, The Samuel Roberts Noble Foundation, Ardmore, OK

Accurately modelling and reconstructing genome-scale biological networks that play critical roles in many fundamental functions in plants, such as signaling transduction, metabolism, and gene expression regulation, is essential toward the understanding of gene function, interaction, and cellular behavior at the genome level. Yet, reverse-engineering of genome-scale biological networks and subsequently discovering functional modules remain a very computationally challenging task, which generally requires effective statistical approaches, innovative computational algorithms and efficient algorithm implementations. To infer, modeling and analyze complex heterogeneous biological networks, we developed a series of innovative bioinformatics methods and systems.

Network inference: We present 1) GPLEXUS (http://plantgrn.noble.org/GPLEXUS/) and 2) DeGNServer (http://plantgrn.noble.org/DeGNServer/), both of which integrate a series of innovative algorithms that have been implemented using parallel-computing techniques to accurately reconstruct and analyze genome-wide plant gene-gene association networks from medium- to large-scale expression profile data. The GPLEXUS adopts an ultra-fast estimation for pairwise mutual information (MI) computing while the DeGNServer extends the widely used Context Likelihood of Relatedness (CLR) framework by the integration of six proven correlation or association analysis methods, e.g. *Spearman rank correlation, Pearson correlation, Mutual-information, Maximum information coefficient, Kendall rank correlation, Thei-sen Estimator*; and both servers implement the algorithms in parallel and distributed environment to enable genome-scale gene-gene association inference at extraordinary speed while maintaining high prediction accuracy.

Network modeling, integration and graph-search empowered network analysis: We present HRGRN (http://plantgrn.noble.org/hrgrn/), which is a “graph search”-empowered integrative bioinformatics platform to host, analyze and discover interactions among genes, proteins, compounds, and small RNAs involving in signaling transduction, metabolic, and gene regulatory networks in *Arabidopsis thaliana*.

Subnetwork/functional module discovery: Much of cells’ functions are organized as networking of interacting modules or sub-networks. We present a novel approach to mine functional modules in heterogonous biological networks using Multiplex PageRank method. Starting from several seed genes, we successfully identified an immune-related core function module involved in plant defensive response to pathogens in *A. thaliana* through integrative analysis of gene co-expression-based association networks or protein-protein interaction networks. Further gene set enrichment analysis and literature analysis on this core module validates that our approach is effective and highly promising to mine functional modules in large-scale heterogeneous biological networks.

W932: Systems Genomics

Re-Designing Regulatory Networks Underlying Plant Stress Responses

Katherine Denby[1], Krzysztof Polanski[1], Chris Penfold[2], Iulia Gherman[1] and Presta Consortium[3], (1)University of Warwick, Coventry, United Kingdom, (2)University of Cambridge, Cambridge, United Kingdom, (3)http://warwick.ac.uk/presta, Coventry, United Kingdom

PRESTA Consortium (warwick.ac.uk/presta) [1, 2], Krzysztof Polanski[2], Chris Penfold[2], Iulia Gherman[1, 3] and Katherine Denby[1, 2]

[1] School of Life Sciences, University of Warwick

[2] Warwick Systems Biology Centre, University of Warwick

[3] Synthetic Biology Centre for Doctoral Training, University of Warwick

Plant responses to abiotic and biotic stress involve large-scale transcriptional reprogramming. We are elucidating the gene regulatory networks underlying these transcriptional responses to the environment using a combination of experimental and computational/mathematical tools. We generated high-resolution time series expression data from Arabidopsis leaves following pathogen infection (bacterial and fungal), drought, and high light. This time series data has enabled us to resolve the chronology of these stress responses and identify transient changes in gene expression. We have generated transcriptional network models predicting regulatory relationships between differentially expressed transcription factors and used network features to identify key regulators of Arabidopsis stress responses. Network comparison has highlighted interactions common to multiple responses and those that appear to be more stress-specific. We have extended the network models by identifying groups of genes co-regulated across multiple stress responses and validated the models by experimentally testing regulatory predictions from simulations. Simulations of the network models can be used to predict how to enhance beneficial transcriptional responses to stress and as such increase the stress tolerance of Arabidopsis.

W933: Systems Genomics

Molecular Mechanisms Regulating Heterosis: Systems Analysis of 981 ZmHET Genes Controlling Grain Yield Heterosis in Maize, Zea mays L.

Meiping Zhang, Yun-Hua Liu, Yang Zhang, Hui Zhi, Wenwei Xu, Seth Murray and **Hong-Bin Zhang**, Texas A&M University, College Station, TX

Heterosis or hybrid vigor is one of the most economically important biological phenomenon. Use of heterosis in agriculture helps feed and clothe billions of people in the world. The utilization of heterosis has not only led to dramatic yield increases in crops, including maize, rice, sorghum and other crops, but also dramatically enhanced crop adaptation to climate change and abiotic and biotic stresses relative to non-hybrid varieties,

due to its significantly increased plant vigor. Nevertheless, the molecular mechanisms underlying heterosis remain unclear. Shortage of such knowledge is significantly limiting our understanding and the most effective uses of heterosis in agriculture. We have cloned 981 genes controlling maize grain yield heterosis, named *ZmHET*, using a new gene cloning system. All of the genes have been validated and their contributions to grain yield heterosis have been determined by turning on or up-regulating each of the genes in the genetic backgrounds of eight maize F_1 hybrid and inbred parent combinations. Each of the *ZmHET* genes was found to contribute to grain yield mid-parent heterosis by 32.5% to 252.1%. However, 128 (13.0%) of them resulted in increased heterosis while 853 (87.0%) led to decreased heterosis when turned on or up-regulated. Moreover, we have analyzed the 981 *ZmHET* genes in a systematic manner and the results have revealed several new and exciting discoveries on the molecular mechanisms regulating grain yield heterosis in maize. This provides new concepts, knowledge and strategies to effectively use and manipulate grain yield heterosis in crop breeding and production.

W934: Systems Genomics
Questions and Discussion
Meiping Zhang, Texas A&M University, College Station, TX

W935: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Teaching Bioinformatics Through Developing Case Studies: Elementary School Through College
Joann Mudge, National Center for Genome Resources (NCGR), Santa Fe, NM
Joann Mudge, Anitha Sundararajan, Ingrid Lindquist, Madeline Kwicklis, Gabriella DeFrancesco, Drew Lighthall, Natasha Farmer, Meghan Hill, Katelynn James, Alicia White, Krista D Glazewski, Michele Shuster
Next generation sequencing technologies have revolutionized biology, becoming a ubiquitous tool. Today's biologists, regardless of expertise, need to be able to convert large sequence datasets into meaningful biological data using bioinformatics. Yet many biology students have had little or no exposure to bioinformatics.
We address this by developing, presenting, and distributing computer-based bioinformatics teaching case studies in the form of engaging scientific stories or puzzles, tailored and implemented from 3rd grade through college. Many have been developed or expanded in collaboration with high school students or elementary and middle school teachers.
Most are self-contained activities for a single class period, but many have scaffolding activities that allow them to serve as more extensive science units in a classroom. Implementation requires only web access and a browser, and they have been successfully completed on desktop and laptop computers, chromebooks, ipads, tablets and smartphones. All case studies use DNA sequences retrieved from Genbank and several are based on scientific papers.
Case study development is itself a teaching tool, immersing the student and/or instructor) in the nuances of sequencing and bioinformatics, teaching them how to troubleshoot, and allowing them to dig deeply into a topic of their choosing. The experience ultimately provides a classroom case study that brings bioinformatics to a much broader audience. By engaging students and teachers in the earliest grade levels, we are providing a solid scientific foundation for students to build on, as well as providing an interesting framework that may spark an enduring interest in STEM fields.

W936: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Bioinformatics Tutorials leveraging the BioExtract Server
Carol Lushbough, University of South Dakota, Vermillion, SD
The BioExtract Server (bioextract.org) is a Web-based system designed to aid researchers in the analysis of distributed life science data using distributed analytic tools by providing a platform to facilitate the creation of bioinformatic workflows. Scientific workflows are created within the system initially by recording the task sequence performed by the user. These tasks may include querying multiple data sources, saving query results as searchable data extracts, and executing local and Web-accessible analytic tools. The series of recorded tasks can then optionally be saved as a reproducible workflow and is available for subsequent re-execution with the same or modified input. In order to handle the vast quantities of biological data generated by high-throughput experimental technologies, the BioExtract Server (bioextract.org) has leveraged iPlant Collaborative (www.iplantcollaborative.org) functionality through their AGAVE REST API to help address big data storage and analysis issues in the bioinformatics field. Leveraging the iPlant cyberinfrastructure has several key advantages: 1) it provides the ability to easily manage, analysis, and share big data, 2) it provides large scale computation support, 3) it is open for the community to contribute new analytic applications, 4) it allows users to easily integrate their analytic tools with other popular application in the development of automated workflows. The integration of iPlant resources through their API has opened up the ability for researchers to easily integrate their own analytic tools into the BioExtract Server, allowing them to be executed on a HPC platform, shared with collaborators, and included in analytic workflows. Because of its ease-of-use, the BioExtract Server offers a great platform for Bioinformatics instruction.

W937: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics
Modified Moore Method for an Undergraduate Bioinformatics Survey Course
John Hsieh, Iowa State University, Ames, IA
The Moore Method was originally developed by R.L. Moore to teach advanced mathematics in the college setting. There have been many adaptations of the Moore Method, under the broad term Modified Moore Method (M^3), which are now classified as a variant of inquiry based learning (IBL). Despite the growing popularity of M^3, it is rarely applied beyond mathematics. At Iowa State University, we designed and taught an "Introduction to Bioinformatics" survey course using M^3 for the first time during Fall semester 2015. The class size was small (n=12), and students all had a background in the natural sciences, most in the biological sciences. Students had little to no formal training in computational sciences. During the 16-week course, students learned to: 1) work on a remote Linux server, 2) read and write Python code, 3) tackle classic bioinformatics problems, and 4) solve current bioinformatics problems with available tools. As with all M^3 courses, learning objectives were met through carefully designed questions given to students prior to each class session. Class sessions were completely led by students (i.e., reversed classroom) presenting solution to the assigned questions. The application of M^3 to our course has led to several desirable

student outcomes: 1) engagement and ownership of the course material, 2) development of a strong sense of community, and 3) uniform learning outcomes. One of the difficulties we experienced with applying M^3 was the creation of the course material. It was tough to create questions that were challenging enough without overwhelming the students.

W938: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics

Recognizing Mutants Among Us: Helping Students Understand the Connections Between Genotype and Phenotype

Jelena Brkljacic, The Arabidopsis Biological Resource Center, Center for Applied Sciences, The Ohio State University, Columbus, OH

The Arabidopsis Biological Resource Center (ABRC) has taken a lead in the design and the assembly of educational resources focused on plants (http://abrcoutreach.osu.edu). Following the development of the Greening the Classroom program for K-12 instruction and the release of a number of college-level teaching modules supported by the American Society of Plant Biologists (ASPB), ABRC is initiating a new citizen science program with the goal to help students make connections between genotypes and phenotypes. This multidisciplinary program will recruit students who will engage in collecting specific phenotypic data on a number of mutant lines from the SALK Institute-generated homozygous sequence-indexed T-DNA insertion collection, for which the positional information about a mutation is known. The data will be compiled in a dedicated website and analyzed further to correlate the known mutant genotype with a phenotype identified by students. Compiled data will be easily searchable for a genotype or a phenotype of interest, enabling researchers to reveal previously unknown connections between mutation in genes and corresponding phenotypes. By phenotyping previously uncharacterized mutant lines, students will actively participate in scientific research, generating large datasets and driving research forward. Students will also gain a better understanding of the consequences associated with the genetic makeup of an organism that might affect development and/or performance. The expected educational impact of this project goes beyond analyzing plant phenotypes, as plants will be used only as a proxy to make general conclusion about the genotype-phenotype links, including those impacting human health.

W939: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics

Connecting Genotype to Phenotype in 7-12 Classrooms with iTAG Barley

Roger Wise, Corn Insects and Crop Genetics Research, USDA-Agricultural Research Service, Iowa State University, Ames, IA, Laurie McGhee, Colfax-Mingo High School, Colfax, IA, Nick Hayes, Cedar Rapids Kennedy High School, Cedar Rapids, IA, Ron Schuck, Ames High School, Ames, IA, Julie Gonzalez, Des Moines Area Community College, Ankeny, IA, Lance Maffin, Bondurant-Farrar High School, Bondurant, IA, Taylor Hubbard, Ankeny High School, Ankeny, IA, Garrett Hall, Southeast Polk High School, Pleasant Hill, IA, Ehren Whigham, Iowa Western Community College, Council Bluffs, IA and Greg Fuerst, USDA-ARS, Iowa State University, Ames, IA

iTAG "Inheritance of Traits and Genes" is a NSF-sponsored Research Experience for Teachers (RET) to promote understanding of the relationship between genotype and phenotype, which is the core foundation for modern genomics projects. Using the diverse Oregon Wolfe Barley population as the model, we have created a self-sustaining, inquiry-based curriculum comprising lab and classroom activities for high school students to learn concepts in plant development, phenotypic diversity, genetics, and/or genomics. Three key systems are presented; homeotic mutations, domestication, and epistasis. Through inquiry based learning, students become better-informed citizen scientists. **iTAG Barley** is aligned to the National Science Standards, and thus, can be adapted to any state standards. Teachers and workshop participants conduct pre- and post content-based survey of **iTAG** concepts. Results of the assessment are incorporated into publications and presented at NSTA and ASPB conferences.

The curriculum for iTAG Barley is available as teacher and student versions [PDF or digital textbook (**iTAG for iPAD;** https://itunes.apple.com/us/book/itag-barley-9-12-curriculum/id959451733?mt=11), and includes NSF-funded thermal cyclers, microcentrifuges, gel boxes, transilluminators, pipetteman, and reagents. It has been implemented in >35 high-school classrooms from 2009-2014, impacting >1,000 students, half of which were underrepresented from urban to rural communities. The project is continuing to expand its reach with the first iTAG Barley workshop hosted by Iowa State University July 28th-31st, 2015. Workshop organizers and participants collectively will use iTAG Barley in 53 classes during the 2015-16 school year, impacting an additional 1,400 high school students. Additional workshops for summer 2016 are already being planned at Iowa State University in Ames IA and Tuskegee University in Tuskegee, AL.

Supported by the National Science Foundation - Plant Genome grants 09-22746 and 13-39348.

W940: Teaching Genetics, Genomics, Biotechnology, and Bioinformatics

Teaching Bioinformatics Data Analysis Using Cloud Computing Technology

Vivek Krishnakumar[1], Haibao Tang[2], Shelby L. Bidwell[1], Benjamin D. Rosen[3], Yongwook Choi[1], Agnes P Chan[1] and Christopher D. Town[1], (1)J. Craig Venter Institute, Rockville, MD, (2)University of Arizona, Tucson, AZ, (3)ARS, USDA, Beltsville, MD

Genome and transcriptome sequencing have become commonplace in labs all over the world. As next generation sequencing technologies emerge, evolve, and become more affordable, researchers need to be abreast of new methodologies and tools to efficiently analyze vast amounts of data generated by these sequencing platforms.

As part of the outreach activities of the NSF-funded Plant Genome Research Resource project targeted towards curating the Medicago truncatula genome, the JCVI Plant Genomics group has been involved in organizing an annual week-long Plant Bioinformatics Workshop (started in 2009, concluded in 2014), open to researchers within and outside the USA.

The workshop covers a wide variety of topics, starting off with the basics of command-line unix for bioinformatics, and progressing on to the bioinformatics processes and analyses used to sequence, assemble and annotate a eukaryotic genome, using Medicago truncatula as a use case example. Guest instructors are invited to present on domain specific topics such as repeat analysis, small RNAs, comparative genomics, etc.

For the hands-on parts of the workshop, in-person participants (limited to 20) utilized JCVI computational resources to perform all the data analysis. Remote attendee presence was managed by WebEx conferencing, allowing for interactive audience participation. Compute resources for these participants was allocated on the Amazon Elastic Cloud Compute (EC2) infrastructure.

All workshop materials are accessible to the public via a virtual machine (VM) image hosted on the iPlant Atmosphere cloud computing environment (https://atmo.iplantcollaborative.org/application/images/899) along with the presentations and handouts on Google Drive (http://j.mp/jcvi-bioinfo-workshop). Interested users can spawn an instance of the VM image and work through the exercises at their own pace.

W941: The Analysis and Role of the Microbiome

Assembling Whole Genomes from Mixed Microbial Communities Using Hi-C

Ivan Liachko, University of Washington - Department of Genome Sciences, Seattle, WA

Assembly of whole genomes from next-generation sequencing is inhibited by the lack of contiguity information in short-read sequencing. This limitation also impedes metagenome assembly, since one cannot tell which sequences originate from the same species within a population. We have overcome these bottlenecks by adapting a chromosome conformation capture technique (Hi-C) for the deconvolution of metagenomes and the scaffolding of *de novo* assemblies of individual genomes.

In modeling the 3D structure of a genome, chromosome conformation capture techniques such as Hi-C are used to measure long-range interactions of DNA molecules in physical space. These tools employ crosslinking of chromatin in intact cells followed by intra-molecular ligation, joining DNA fragments that were physically nearby at the time of crosslink. Subsequent deep sequencing of these DNA junctions generates a genome-wide contact probability map that allows the 3D modeling of genomic conformation within a cell. The strong enrichment in Hi-C signal between genetically neighboring loci allows the scaffolding of entire chromosomes from fragmented draft assemblies. Hi-C signal also preserves the cellular origin of each DNA fragment and its interacting partner, allowing for deconvolution and assembly of multi-chromosome genomes from a mixed population of organisms.

We have used Hi-C to scaffold whole genomes of animals, plants, fungi, as well as prokaryotes and archaea. We have also been able to use this data to annotate functional features of microbial genomes, such as centromeres in many fungal species. Additionally, we have applied our technology to diverse metagenomic populations such as craft beer, bacterial vaginosis infections, soil, and tree endophyte samples to discover and assemble the genomes of novel strains of known species as well as novel prokaryotes and eukaryotes. The high quality of Hi-C-based assemblies allows the simultaneous closing of numerous unculturable genomes, placement of plasmids within host genomes, and microbial strain deconvolution in a way not possible with other methods.

W942: The Analysis and Role of the Microbiome

Discovery and Annotation of Novel Proteins from Rumen Gut Metagenomic Sequencing Data

Mick Watson, The Roslin Institute and R(D)SVS, University of Edinburgh, Edinburgh, United Kingdom

Metagenomics is the practice of sequencing all genomes within an ecosystem, and has become the gold standard for the study of microbial population and structure in environmental samples. However, to make the most of metagenomic data requires a suite of sophisticated bioinformatics software. One particular focus of metagenomics has been the identification of novel enzymes and other proteins with functions relevant to the biotechnology and pharmaceutical industries. As more technologies, datasets, software, and analytical systems are invented, easy access, analysis, evaluation and sharing of the data are required. Here I review tools and pipelines for the discovery and annotation of novel proteins from metagenomics data, and the results of applying those pipelines to a number of rumen samples from Scotland

W943: The Analysis and Role of the Microbiome

TBD

Rob Knight, University of California at San Diego, La Jolla, CA

W944: The Analysis and Role of the Microbiome

TBD

Scot Hulbert, Washington State University, Pullman, WA

W945: The Phytoremediation Genome

Endophyte-Assisted Phytoremediation of TCE, PAHs, TNT, and Arsenic

Sharon L. Doty[1], Zareen Khan[1], Jun Won Kang[2], Andrea Firrincieli[3], Robert Tournay[1], Dominic Sivitilli[1], Mitch K. Scott[1], John L. Freeman[4] and Michael J. Blaylock[5], (1)University of Washington, Seattle, WA, (2)Korea Forest Research Institute, Suwon, South Korea, (3)University of Tuscia, Viterbo, Italy, (4)Intrinsyx Technologies Corporation, Moffett Field, CA, (5)Edenspace Systems Corporation, Purcellville, VA

Phytoremediation is an effective technology for remediation of a wide range of environmental pollutants. However, when the pollutant is at phytotoxic concentrations, a modified technology is required. Endophytes are microorganisms that live within plants that can provide multiple benefits to the plant including increased nutrient acquisition and growth, and improved stress tolerance. In addition, some strains of endophytes have the capacity to degrade or otherwise reduce the toxicity of xenobiotics. We have isolated endophytes that degrade TCE, PAHs, and TNT, and addition of the strains to plants results in improved plant health on normally phytotoxic levels of these pollutants. The genomes of the TCE and PAH degrading strains have recently been sequenced, providing insight into mechanisms by which the strains can improve phytoremediation. Using fluorescent microscopy, we have demonstrated that multiple strains can co-colonize poplar, providing the potential to equip plants for sites polluted with multiple classes of chemicals. A field trial of the TCE-degrading strain with poplar trees is currently underway. In addition to the endophytes capable of degrading organic pollutants, we have recently isolated an arsenic-tolerant endophyte that reduces the phytotoxic effects of arsenic in inoculated Arabidopsis seedlings. Endophyte-assisted phytoremediation is an important technological advance enabling this green technology to be utilized on a broader spectrum of contaminated areas.

W946: The Phytoremediation Genome

Using Elemental Profiling and Systems Biology to Identify Genes Underlying Toxic Element Uptake in Plants

Ivan Baxter, USDA-ARS Plant Genetics Research Unit, St. Louis, MO; Danforth Plant Science Center, St. Louis, MO
The vast majority of the elements that make up a seed, with the exception of carbon and oxygen, are obtained from soil via the roots. These soil-derived elements are required for plant structure, metabolism, protein function, signaling, and proper osmotic and electrostatic potential. Elemental accumulation requires the integration of processes across biological scales, including interactions with the soil matrix and biota, subcellular localization, metabolism, and gas exchange. Thus, the elemental composition of seeds (the "ionome"), including both beneficial and toxic elements, is a consequence of complex plant-environment interactions with serious nutritional implications. High-throughput ionomics workflows allow a single inductively coupled plasma mass spectrometer (ICP-MS) to precisely analyze hundreds of samples for more than 20 elements per day. We have used ionomic profiling of 200,000+ maize kernels, soybeans and cotton seeds to detect the genetic and environmental determinants of the ionome. Using modern genetic approaches, we can easily identify loci that control the ionome in a given environment. We are using several different approaches to integrate ionomic and other systems biology datasets to identify the causal genes underlying these phenotypes. We can use this knowledge to identify genes important for toxic element accumulation in plants.

W947: The Phytoremediation Genome

Metal Hyperaccumulators

Ute Kraemer[1], Terezie Mandakova[2], Vasantika Singh[1], Ricardo J. Stein[1], Stephan Hoereth[3], J. Romario F. de Melo[1], Aitor F. Gonzaga-Moltó[1], Felix Mathias Bemm[4], Marc Hanikenne[5], Detlef Weigel[6], Stephan Clemens[3] and Martin A. Lysak[2], (1)Department of Plant Physiology, Ruhr University Bochum, Bochum, Germany, (2)CEITEC, Masaryk University, Brno, Czech Republic, (3)University of Bayreuth, Bayreuth, Germany, (4)Max Planck Institute for Developmental Biology, Department Weigel, Genome Informatics, Tuebingen, Germany, (5)University of Liège, Liège, Belgium, (6)Max Planck Institute for Developmental Biology, Tuebingen, Germany

A metal hyperaccumulator is defined as a plant species of which any one individual, when grown in its natural habitat in the wild, contains an extraordinarily high level of a metal or metalloid in above-ground tissues of > 10,000 µg g^{-1} Mn, > 3,000 µg g^{-1} Zn, > 1,000 µg g^{-1} Ni, Se, As, Co, Pb, Cu, Th, or Sb, or > 100 µg g^{-1} Cd in dry biomass. These concentrations are more than an order of magnitude above the critical toxicity thresholds in ordinary plants. Thus, metal hyperaccumulation requires extraordinarily efficient and specific mechanisms of metal extraction from the soil, root-to-shoot metal partitioning, and internal metal detoxification. Metal hyperaccumulation and associated hypertolerance in plants bear great promise for the development of phytoremediation and phytomining technologies, in addition to serving as extreme model traits in molecular physiology, evolution and ecology. About 50% of the ~500 known metal hyperaccumulator taxa are in the Brassicaceae family, of which two species – the Zn/Cd/Ni hyperaccumulator *Noccaea caerulescens* and the Zn/Cd hyperaccumulator *Arabidopsis halleri* – have emerged as new model species. Various genome-wide comparative approaches including closely related non-hyperaccumulator species have yielded important insights into both molecular mechanisms and the evolution of metal hyperaccumulation. Through reverse genetics, we have demonstrated the functions of key metal hyperaccumulation genes. We are presently assessing species-wide phenotypic diversity in *A. halleri*, and taking first steps towards its genetic analysis. The current status of this work will be presented in relation to traits of interest for the development of phytoremediation technologies.

W948: The Phytoremediation Genome

Systems Level Approaches Towards Understanding Heavy Metal Response and Resistance in Plants

Julian I. Schroeder, Andrew M Cooper, Felix Hauser, Qingqing Xie and Tim O Jobe, University of California, San Diego, La Jolla, CA

Uptake of toxic heavy metals and metalloids in plants primarily occurs through transporters designed to import essential nutrients. Cadmium uptake is mediated by zinc and iron (ZIP) transporters and arsenic by organic phosphate transporters. Therefore plants have developed a complex and dynamic system to detoxify and sequester heavy metals to minimize detrimental effects. Exposure to heavy metals causes rapid and diverse changes in gene expression to control and initiate the detoxification machinery. While several key mechanisms in the chelation and transport pathways have been identified, the signaling network behind this rapid gene regulation remains largely unknown. We are pursuing two distinct methods In order to uncover components of the Cd-specific transcriptional response network. First, using our Cd-dependent transcriptome data sets, we generated an Arabidopsis line carrying a cadmium-inducible promoter (SULTR1;2) fused to a luciferase reporter gene, which was then mutagenized using ethyl methanesulfonate (EMS). Mutants were then screened for shifts in Cd-inducible luciferase responses. Identified mutants were categorized into three groups based on the response: super response to cadmium (SRC), constitutive response to cadmium (CRC), and non-response or reduced response to cadmium (NRC). Two non-response mutants, NRC1 and NRC2, have been mapped and characterized as γ-glutamylcysteine synthetase and glutathione synthetase, respectively, both of which are involved in the generation of chelators essential for heavy metal detoxification. The NRC2 mutation is the first viable recessive mutation in the glutathione synthetase gene. We are currently working to map and characterize two additional mutants, one super response mutant, SRC1, and one constitutive response mutant, CRC1, using segregation and genomic resequencing.

Traditional forward genetic screens in plants are limited in the identification of homologous genes with overlapping functions, with only approximately 10% of Arabidopsis genes having been linked to a single-gene mutant phenotype. Thus (partial) genetic redundancies greatly limit forward genetic screens. To address over-lapping gene functions on a genomic systems level scale, the P.I.'s laboratory has computationally designed artificial microRNA (amiRNA) libraries using over 90,000 hours CPU time at the San Diego Supercomputer Center, for genome-wide knock-down of homologous gene family members. We generated 10 amiRNA libraries consisting of 22,000 amiRNAs in total, with 96% of the amiRNAs targeting 2 to 5 genes. This enables rapid progression from phenotype to gene. System level screens of amiRNA plants grown in the presence of cadmium or arsenic are identifying a wealth of new genes that affect these toxic heavy metal and metalloid responses.

W949: The Phytoremediation Genome

TBD

Norman Terry, Department of Plant and Microbial Biology, Berkeley, CA

W950: The Resurgence of Reference Quality Genome Sequence

The Resurgence of Reference Quality Genomes

Michael Schatz[1], Hayan Lee[2], James Gurtowski[1], Shinjae Yoo[3], Maria Nattestad[1], Shoshana Marcus[4], Sara Goodwin[1] and W. Richard McCombie[1], (1)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (2)Stony Brook University, Stony Brook, NY, (3)Brookhaven National Laboratory, Upton, NY, (4)City University of New York, Brooklyn, NY

Several new 3rd generation long-range DNA sequencing and mapping technologies have recently become available that are starting to create a resurgence in genome sequence quality. Unlike their 2nd generation, short-read counterparts that can resolve a few hundred or a few thousand base-pairs, the new technologies can routinely sequence 10,000 bp reads or map across 100,000 bp molecules. The substantially greater lengths are being used to enhance a number of important problems in genomics and medicine, including de novo genome assembly, structural variation detection, and haplotype phasing.

Here we discuss the capabilities of the latest technologies, and show how they will improve the “3Cs of Genome Assembly”: the contiguity, completeness, and correctness. We derive this analysis from (1) a meta-analysis of the currently available 3rd generation genome assemblies, (2) a retrospective analysis of the evolution of the reference human genome, and (3) extensive simulations with dozens of species across the tree of life.

Overall, we anticipate these technologies will unlock the genomic “dark matter”, and provide many new insights into evolution, agriculture, and disease.

W951: The Resurgence of Reference Quality Genome Sequence

High Quality, Highly Contiguous Genome Assemblies Now

Nicholas H. Putnam[1], Brendan O'Connell[1], Jonathan C. Stites[1], Brandon J. Rice[1], Andrew Fields[1], Paul D. Hartley[1], Charles W. Sugnet[1], David Haussler[2], Daniel S. Rokhsar[3] and **Richard E. Green**[1,2], (1)Dovetail Genomics, Santa Cruz, CA, (2)UC Santa Cruz, Santa Cruz, CA, (3)DOE Joint Genome Institute, Walnut Creek, CA

We have developed an efficient approach for high-quality genome assembly using only generic high-throughput Illumina sequence data. The key to our approach is generation of a Chicago library, via in vitro proximity ligation, that simultaneously reveals genome contiguity at a range of distances up to the size of input DNA. Using these data and a sophisticated genome assembly pipeline called HiRise, we have generated dozens of highly contiguous, highly accurate genome assembles from a wide variety of genomes from across the plant and animal kingdoms. This streamlined approach reduces the total time for de novo genome assembly to under 10 weeks. I will highlight recent improvements in our technique including reduced input requirements, gap-filling, quality assessment, and highlight the biological discovery possible when genomes are assembled at multi-megabase scaffold N50s which we routinely achieve.

W952: The Resurgence of Reference Quality Genome Sequence

Scalable Parallel Algorithms for *de novo* Assembly of Complex Genomes

Evangelos Georganas[1], **Aydin Buluc**[1], Jarrod Chapman[2], Steven Hofmeyr[1], Rob Egan[3], Chaitanya Aluru[4], Leonid Oliker[1], Daniel S. Rokhsar[2] and Katherine Yelick[1], (1)Lawrence Berkeley National Laboratory, Berkeley, CA, (2)DOE Joint Genome Institute, Walnut Creek, CA, (3)DOE Joint Genome Institue, Walnut Creek, CA, (4)UC Berkeley, Berkeley, CA

De novo whole genome assembly reconstructs genomic sequence from short, overlapping, and potentially erroneous DNA segments and is one of the most important computations in modern genomics. This work presents the first high-quality end-to-end de novo assembler designed for extreme scale analysis, via efficient parallelization of the Meraculous code. In this talk, I will first present distributed-memory parallelization of de Bruijn graph construction and traversal, which is a key component of most de novo genome assemblers. I will also talk about load-balancing techniques for repetitive genomes with highly skewed k-mer distributions. Then, I will briefly talk about merAligner, a highly parallel sequence aligner that implements a seed-and-extend algorithm. Since merAligner employs parallelism in all of its components, especially the seed index construction, it is particularly useful for aligning contigs to reads within the context of de novo genome assembly. Finally, I will mention parallelization of the Meraculous scaffolding modules by leveraging the one-sided communication capabilities of the Unified Parallel C (UPC) while effectively mitigating load imbalance. Large-scale results on a Cray XC30 using grand-challenge genomes demonstrate efficient performance and scalability on thousands of cores. Overall, our pipeline accelerates Meraculous performance by orders of magnitude, creating unprecedented capability for extreme-scale genomic analysis.

W953: The Resurgence of Reference Quality Genome Sequence

Using PacBio Long Reads to Generate a High Quality Reference for the Allotetraploid *Coffea arabica* and its Maternal Diploid Ancestor *Coffea eugenioides*

Marcela Yepes[1], Alvaro Gaitan[2], Marco A. Cristancho[3], Luis Fernando Rivera[3], Juan Carlos Correa[3], Carlos Ernesto Maldonado[2], Carmenza E. Gongora[2], Andres Mauricio Villegas[2], Huver Posada[4], Aleksey Zimin[5], James A Yorke[5], Keithanne Mockaitis[6] and Herb Aldwinckle[1], (1)Cornell University/ School of Integrative Plant Sciences/ Plant Pathology and Plant Microbe Biology Section, Geneva, NY, (2)Centro Nacional de Investigaciones de Cafe, CENICAFE, Chinchiná, Colombia, (3)Colombian Center for Bioinformatics and Computational Biology (Bios), Manizales, Caldas, Colombia, (4)Federacion Nacional de Cafeteros de Colombia (FNC)/ Centro Nacional de Investigaciones de Café (CENICAFE), Chinchina, Caldas, Colombia, (5)University of Maryland, College Park, MD, (6)Indiana University, Bloomington, IN

Allopolyploids originate from hybridization between divergent genomes associated with chromosome set doubling. As a consequence, the genomes may undergo a wide range of structural, epigenetic, and functional changes. The world’s most widely cultivated coffee species, representing 70% of the coffee market, is the allotetraploid, *Coffea arabica* (2n=4x=44; genome size 1.3 Gb). *C. arabica* evolved through the interspecific hybridization of the ancestors of two diploid *Coffea* species: *Coffea eugenioides* (2n=22, maternal donor, genome size 0.66 Gb) and *C. canephora* (2n=22, paternal donor, genome size 0.71 Gb). Sequencing and assembly of the *C. canephora* genome was published recently, Denoeud *et al.* 2014. Science 345: 1181-1184; genome assembly can be accessed at: http://coffee-genome.org. We report here progress to

produce high quality reference assemblies for *C. eugenioides* and *C. arabica* using Pacific BioSciences (PACBio) long reads to enable coffee genetics and genomics of coffee and speed up adaptation of the crop to climate change. Climate change is probably the most severe threat currently facing the coffee industry on the global scale. In recent years, extreme weather events in Central America, Colombia, and Brazil have led to coffee production losses of more than US $2 bn. Of major concern is the very narrow genetic base of cultivated coffee varieties, and therefore the urgent need to develop advanced genomic tools to speed up characterization of *Coffea* diversity in its Center of Origin, Ethiopia, which accounts for 98% of the genetic pool, to help broaden the genetic base of cultivated *C. arabica* and speed up adaptation of the crop to climate change.
This abstract will be presented by co-authors Marcela Yepes and Marco Cristancho.

W954: The Resurgence of Reference Quality Genome Sequence

MaSuRCA Mega-Reads Assembly Technique for Haplotype Resolved Genome Assembly of Hybrid PacBio and Illumina Data

Aleksey Zimin[1], James A. Yorke[1] and Guillaume Marcais[2], (1)University of Maryland, College Park, MD, (2)Insitutte for Physical Science and Technology, College Park, MD

The developments in DNA sequencing technology over the past several years have enabled large number of scientists to obtain sequences for the genomes of their interest at a fairly low cost. Illumina Sequencing was the dominant whole genome sequencing technology over the past few years due to its low cost. The Illumina reads are short (up to 300bp) and thus most of those draft genomes produced from Illumina data are very fragmented which limits their usability in practical scenarios. Longer reads are needed for more contiguous genomes. Recently Pacbio sequencing made significant advances in developing cost-effective long-read (>10000bp) sequencing technology and their data, although several times more expensive than Illumina, can be used to produce high quality genomes. Pacbio data can be used for de novo assembly, however due to its high error rate high coverage of the genome is required this raising the cost barrier. A solution for cost-effective genomes is to combine Pacbio and Illumina data leveraging the low error rates of the short Illumina reads and the length of the Pacbio reads. We have developed MaSuRCA mega-reads assembler for efficient assembly of hybrid data sets and we demonstrate that it performs well compared to the other published hybrid techniques. Another important benefit of the long reads is their ability to link the haplotype differences. The mega-reads approach corrects each Pacbio read independently and thus haplotype differences are preserved. Thus, leveraging the accuracy of the Illumina data and the length of the Pacbio reads, MaSuRCA mega-reads can produce haplotype-resolved genome assemblies, where each contig has sequence from a single haplotype. We present preliminary results on haplotype-resolved genome assemblies of faux (proof-of-concept) and real data.

W955: The Resurgence of Reference Quality Genome Sequence

How to Compare and Cluster Every Known Genome in about an Hour

Sergey Koren[1], Brian D Ondov[2], Todd J. Treangen[2], Adam B Mallonee[2], Nicholas H. Bergman[2] and Adam M Phillippy[1], (1)National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, (2)National Biodefense Analysis and Countermeasures Center, Frederick, MD

Given a massive collection of sequences, it is infeasible to perform pairwise alignment for basic tasks like sequence clustering and search. To address this problem, we demonstrate that the MinHash technique, first applied to clustering web pages, can be applied to biological sequences with similar effect, and extend this idea to include biologically relevant distance and significance measures. Our new tool, Mash, uses MinHash locality-sensitive hashing to reduce large sequences to a representative sketch and rapidly estimate pairwise distances between genomes or metagenomes. Using Mash, we explored several use cases, including a 5,000-fold size reduction and clustering of all 55,000 NCBI RefSeq genomes in 46 CPU hours. The resulting 93 MB sketch database includes all RefSeq genomes, effectively delineates known species boundaries, reconstructs approximate phylogenies, and can be searched in seconds using assembled genomes or raw sequencing runs from Illumina, Pacific Biosciences, and Oxford Nanopore. For metagenomics, Mash scales to thousands of samples and can replicate Human Microbiome Project and Global Ocean Survey results in a fraction of the time. Other potential applications include any problem where an approximate, global sequence distance is acceptable, e.g. to triage and cluster sequence data, assign species labels to unknown genomes, quickly identify mis- tracked samples, and search massive genomic databases. In addition, the Mash distance metric is based on simple set intersections, which are compatible with homomorphic encryption schemes. To facilitate integration with other software, Mash is implemented as a lightweight C++ toolkit and freely released under a BSD license athttps://github.com/marbl/mash

W956: Transposable Elements

Uncover Hidden DNA Patterns of Helitrons in Plant Genome

Wenwei Xiong[1], Hugo Dooner[2] and **Chunguang Du**[1], (1)Montclair State University, Montclair, NJ, (2)Rutgers University, Piscataway, NJ

Helitrons are unique from other types of transposable elements because they have the capability to capture gene fragments and carry them throughout the genome. However they are difficult to identify because they lack the classic characteristics found in other types of transposable elements. Here we describe HelitronScanner, a generalized tool for their identification based on a motif-extracting algorithm proposed initially in a study of natural languages. We applied a two-layered Local Combinational Variable (LCV) approach for generalized *Helitron* identification. LCV represents any combination pattern of either nucleotides or the associations of these nucleotide patterns. The first layer extracts location-nonspecific LCVs (n-LCV) in a known *Helitron* set and then creates a distribution matrix of these n-LCVs matching against *Helitrons*. The second layer draws location-specific LCVs (s-LCV) from the distribution matrix. This two-layered procedure is applied to new sequences comparably. In HelitronScanner, n-LCVs represent sequence patterns shared by *Helitrons* and s-LCVs constitute the associations of these patterns within *Helitrons*. HelitronScanner overcomes the divergence of *Helitron* termini among species by using conserved nucleotides at potentially variable locations. We ran HelitronScanner against a wide range of plant genome sequences from Phytozome version 9.0 and identified a total of 107,367 *Helitrons*. Many new *Helitrons* were identified in model species, such as maize, rice, and *Arabidopsis*, and in a

variety of organisms where *Helitrons* had not been reported previously, leading to a major upward reassessment of their abundance in plant genomes. *Helitron* abundance varied greatly among sequenced genomes.

W957: Transposable Elements

DNA Transposons Specifically Accelerate the Evolution of Genes in Grasses

Thomas Wicker, University of Zurich, Zurich, Switzerland

W958: Transposable Elements

Centromeric Sequences Change in Wheat and its Distant Hybrids

Handong Su[1], Xiang Guo[2], Qinghua Shi[2], Shulan Fu[3], Jing Wang[2], Xiangqi Zhang[2], Zanming Hu[4] and Fangpu Han[5], (1)Institute of Genetics and Developmental Biology, Beijing, China, (2)Institute of Genetics and Developmental Biology, Beijing, China, (3)Sichuan Agricultural University, Wenjiang, China, (4)Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China, (5)Chinese Academy of Sciences, Beijing, China

Centromeres typically contain repeat sequences, but centromere function does not necessarily depend on these sequences. In aneuploid wheat (*Triticum aestivum*) and the offspring of wide hybrid, we found functional centromeres with severe quantitative changes regarding the content of centromeric retrotransposon of wheat (CRW) sequences. CRWs were strongly reduced in the ditelosomic lines 1BS, 5DS, 5DL, which did not affect centromere function. A wheat-*Thinopyrum elongatum* addition line and they were completely lost in the ditelosomic line 4DS. A 994-kb ectopic genomic DNA sequence near the former centromeres was involved in *de novo* centromere formation on 4DS chromosome. The 994-kb region revealed no differential enrichment of histone modification comparing to normal centromere. In addition, two ectopic sequences were incorporated in a *de novo* centromere of a wheat-*Th. intermedium* addition line. Stable alien chromosomes with two and three regions containing CRW sequences were found in wheat-*Th. elongatum* hybrid derivatives, but only one represented a functional centromere. In wheat-rye (*Secale cereale*) hybrids, rye centromere specific sequences spread along one the chromosome arm and may cause centromere expansion. Thus, Chromosome rearrangement during the formation of wheat aneuploidy and its distant hybrid caused centromere sequence elimination, expansion and multi-centromere formation, which may be associated with chromosome stability and novel chromosome formation.

W959: Transposable Elements

Elucidation of Transposable Elements in Conifers and their Effect on Conifer Evolution

Robin Paul[1], Kristian A. Stevens[2], Daniel Gonzalez-Ibeas[1], Kevin Pratt[1], Aleksey Zimin[3], James A. Yorke[3], Ann Holtz-Morris[4], Maxim Koriabine[4], Marc Crepeau[2], Daniela Puiu[5], Steven L. Salzberg[5], Pieter J. deJong[4], Charles H. Langley[2], David B. Neale[6] and Jill L. Wegrzyn[1], (1)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, (2)Department of Evolution and Ecology, University of California, Davis, Davis, CA, (3)University of Maryland, College Park, MD, (4)Children's Hospital Oakland Research Institute, Oakland, CA, (5)Johns Hopkins University, School of Medicine, Baltimore, MD, (6)Dept. Plant Sciences University of California Davis, Davis, CA

Conifers are the most dominant life forms in boreal and temperate forests. They have important applications towards carbon sequesterization and wood production. Interspersed repeats constitute up to 80% of these typically large and complex genomes (Douglas-fir: 15Gbp, loblolly pine: 23Gbp and sugar pine: 33Gbp) making elucidation of these genomes extremely challenging. A combination of de novo and similarity based methods were used to characterize the repeats in these complex genomes. It was observed that transposable elements constitute about 70%, 74% and 79% of Douglas-fir, loblolly pine and sugar pine genome respectively. As expected, LTR retrotransposons dominated all three genomes and represent 53.67%, 52.53% and 56.08% of Douglas-fir, loblolly pine and sugar pine genomes respectively. To study the proliferation and diverged nature of the LTRs, dating of LTR retrotransposons were carried out by comparing the nucleotide substitutions of their respective LTR regions. Dating results suggest that transposable element activity has been more recent in sugar pine relative to loblolly pine resulting in the expansion of the sugar pine genome. No major bursts in transposon activity were detected and instead a balanced insertion and removal of transposable elements occurred over the past 15 million years. Therefore no single repeat element contributed a major portion of the genome. It was also found that gypsy elements in sugar pine were more recent than those in loblolly pine supporting our observation of a gypsy expansion in sugar pine.

W960: Transposable Elements

Characterization and Evolutionary Analyses of Terminal-Repeat Retrotransposons in Miniature (TRIMs) in Plants

Dongying Gao[1], Yupeng Li[1], Kyung Do Kim[1], Jason Abernathy[1], Rod A. Wing[2] and Scott A. Jackson[1], (1)University of Georgia, Athens, GA, (2)Arizona Genomics Institute, University of Arizona, Tucson, AZ

Terminal-repeat retrotransposons in miniature (TRIMs) represent a unique group of LTR retrotransposons that are extremely small and difficult to identify. Thus far, only a few TRIMs have been characterized, and their evolutionary impact on host genomes is poorly understood.
We combined *de novo* and homology-based methods to annotate TRIMs in sequenced plant genomes. We identified 145 previously undescribed TRIMs, including the first TRIMs in a lycophyte and in non-vascular plants. The majority of the TRIM families were highly conserved and shared within and between plant families. Unlike most LTR retrotransposons, TRIMs are enriched in or near genes. TRIMs were targeted by both 21 and 24 nt small RNAs and frequently found in CG body methylated genes. Importantly, we identified putative autonomous retrotransposons and very recent transpositions of a TRIM element in *O. sativa*. Our data suggested that TRIMs are widely present in plant genomes and may have unique impacts on host genomes. The pipeline was also used to annotate animal genomes and we found TRIMs in the human, mouse and nematode.

W961: Transposable Elements

The *Tgm9*-Induced Indexed Insertional Mutant Collection to Conduct Community-Based Reverse Genetic Studies in Soybean

Madan K. Bhattacharyya[1], Jordan Baumbach[1], Ronan O'Malley[2] and Devinder Sandhu[3], (1)Iowa State University, Ames, IA, (2)The Salk Institute, La Jolla, CA, (3)United States Department of Agriculture - Agricultural Research Services, Riverside, CA
Until now, functional analyses of soybean genes have been very arduous because of the lack of a rapid transformation procedure. Recently identified the active endogenous type II transposable element, *Tgm9*, excises from insertion sites and restores wild-type phenotypes. Thus, this element provides a great promise in cloning soybean genes bypassing challenges associated with soybean transformation. By applying a novel high-throughput technology, we determined *Tgm9*-insertion sites in 5,184 lines. These *Tgm9* insertions are found in all twenty soybean chromosomes showing preferential re-insertion into gene-rich chromosomal arms as compared to gene-poor centromeric regions. Among the 5,184 mutants, 1,542 represent potential *Tgm9*-induced gene-disruption mutants at a rate of 50 to 106 per chromosome. Considering that only 16% of the soybean genome contains genes, this study establishes that a *Tgm9* mutagenesis population is enriched for potential gene-disruption mutants making *Tgm9* an important resource for functional analysis of soybean genes.

W962: Transposable Elements
Sample Sequence Analysis of Grass Genomes Indicates Frequent and Repeated Horizontal Transfer of LTR-Retrotransposons
Minkyu Park[1], P. a. Christin[2], Colin P. Osborne[2] and Jeffrey L. Bennetzen[1], (1)Dept. of Genetics, University of Georgia, Athens, GA, (2)Department of Animal and Plant Sciences Sheffield University, Sheffield, United Kingdom
LTR-retrotransposons are one of the most important components of plant genomes. A few studies have compared their properties across multiple “fully sequenced” genomes, but this requirement for a genome assembly means that only a small number of taxa can currently be investigated. To demonstrate an approach for comprehensive comparative analysis of LTR-retrotransposons across thousands of species, we generated low-depth sample sequence data for 19 previously-uncharacterized genomes from the panicoid grass subfamily. From the analysis, we identified highly dynamic changes in the activity of numerous LTR-retrotransposon families. The three *Ty3/Gypsy* families *Milt*, *Xilon-Diguus*, and *Grande* were found to be particularly variable, exhibiting very different compositions even among closely related species. Numerous cases of lineage-specific activation and extinctions of specific LTR-retrotransposon families were observed. Using 62 public plant genome sequences and the 19 panicoid sample sequences, we investigated the possible horizontal transfer of LTR-retrotransposons. We found that the genus *Oryza* has had many horizontal transfers with the panicoids, including at least 24 separate horizontal transfers involving 11 *Oryza* species and 19 panicoid species. Among the 11 *Oryza* species, *Oryza sativa* indica exhibited an extraordinarily highly frequency of horizontal transfer with the lineage including *Echinochloa haploclada*.

W963: Tripal Database Network and Initiatives
Tripal v2: An Overview
Lacey-Anne Sanderson[1], Stephen P. Ficklin[2], Dorrie Main[2] and Kirstin Bett[1], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)Washington State University, Pullman, WA
Tripal is a freely available, open-source construction toolkit for community-focused, biological data web portals. It is designed to help research communities to collate data in an organized and searchable manner, as well as, to build their community through research news feeds, conferences, and forums. Through its integration with Drupal, Tripal can decrease the cost of development and maitenance by reducing the amount of site-specific programming required. Tripal v2 provides data-specifc content pages (e.g. for genes, stock, etc.), data loaders (e.g. GFF, FASTA, OBO, GAF and support for tab-delimited files), and simple search tools (e.g. genomic features, germplasm, publications, etc.) by default, with site administrators having the ability to both customize or turn-off functionality as desired. Many customizations can be made through graphical user interfaces including changing page URLs and titles, search filters and results, and even administrative listings. Furthermore, both Drupal and Tripal provide extensive Application Program Interfaces (APIs), which ensure that all aspects of the portal can be tailored to both the community it serves and the data available. Tripal is available for download from http://drupal.org/project/tripal and has extensive documentation available on http://tripal.info.

W964: Tripal Database Network and Initiatives
The Future of Tripal: intuitive content creation, flexible data storage and web services
Stephen P. Ficklin[1], Lacey-Anne Sanderson[2], Chun-Huai Cheng[1], Connor Wytko[1], Brian Soto[1], Mark Clytus[1], Kirstin Bett[2] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)University of Saskatchewan, Saskatoon, SK, Canada
An increasing number of research communities are choosing to use Tripal to publish their genomic, genetic and related data in an online searchable format. Tripal uses the GMOD Chado database schema, a community-derived relational database schema, and Drupal, a popular content management system (CMS), to provide data-specific content pages, data loaders and simple search tools. With the increase in data magnitude, the need for more flexible data storage including direct use of BAM/VCF files and no-SQL technologies has arisen. Furthermore, the desire to exchange and share data has intensified with the need to reduce data duplication. In response to these needs, Tripal version 3 has changed its organizational focus from chado to controlled vocabularies and adopted Drupal Entities. This affords easier integration of non-Chado storage platforms, facilitates implementation of RESTful web services, provides greater search and display functionality and drives more intuitive content creation. Tripal 3 is also designed to require less programming than previous versions when layout customizations are desired.

W965: Tripal Database Network and Initiatives
Overview of NRSP and Planned Tripal development
Dorrie Main, Washington State University, Pullman, WA

W966: Tripal Database Network and Initiatives
Extending Tripal to Manage Banana Genetic Resource Information

Valentin Guignon[1], Alexis Andrieu[1], Stephen P. Ficklin[2], Max Ruas[1], Nicolas Roux[1] and Mathieu Rouard[1], (1)Bioversity International, Montpellier, France, (2)Washington State University, Pullman, WA

Large-scale genetic and phenotypic characterization of germplasm collections has the great potential to change the way scientists deal with genetic resources. Users of genebanks should be in a position to select germplasm material based on a combination of passport, genotypic and phenotypic information among the global genepool. A new generation of information systems has to be designed to efficiently handle this information and link it with others external resources such as genome or breeding databases. The Musa Germplasm Information System (MGIS) addresses the management of banana genetic resources. It is implemented with Drupal content management system using the Tripal module to work with Chado database schema. However, this stack does not address completely all of our requirements. It should allow us to handle access restriction, germplasm ordering, workflows and should track all changes for quality control. Finally, there is a need for interoperability among external resources or fieldbooks. Fortunately, Drupal and Tripal have been designed to support custom extensions like the ones we created. Firstly, the Chado Controller extension manages Chado data access levels, modification history and data integrity checks. The Multi-Chado extension brings more control over data access and enables the use of other Chado instances on a same Tripal instance. Then, the Breeding API extension enables data exchange with external systems. Finally, an additional in-house extension called MGIS module gives us even more flexibility. Based on these examples, we will illustrate how bioinformatics developers can expand Tripal for their needs.

W967: Tripal Database Network and Initiatives

Efforts by the Hardwood Genomics Database team to extend Tripal functionality

Nate Henry, Ming Chen and Margaret Staton, University of Tennessee, Knoxville, TN

The Hardwood Genomics Database team at the University of Tennessee is working to extend the functionality of Tripal by (1) providing a method to easily load and display variable gene expression data in Chado and (2) providing intuitive search tools utilizing the Elasticsearch search engine. Both solutions will be implemented as modules that can be installed in current Tripal websites.

The gene expression module is being developed in an effort to create a standard way to store and display variable expression data in chado, including experimental data derived from RNAseq. In order to be compatible with current Chado databases, tables from the Chado MAGE module were chosen as a method for storing expression data. Tripal data loaders will be provided for the most common expression output formats, and the expression data across different samples (tissues, treatments or other) will be displayed on the feature page.

The Tripal Elasticsearch module is intended to be a replacement for the standard Drupal search functionality as well as an improvement on the search functionality provided by Drupal Views. After implementing the Elasticsearch server software, the Tripal Elasticsearch module will enable the indexing of Drupal nodes as well as all records in the Chado database. The search interface for users will allow a broad site-wide search as well as searches customizable by organism, data type or other limits.

Both the Tripal gene expression module and the Tripal Elasticsearch module will be made publicly available along with a manual to guide implementation.

W968: Tripal Database Network and Initiatives

TripalBIMS: the Breeding Information Management System in Tripal

Sook Jung[1], Taein Lee[1], Chun-Huai Cheng[1], Jing Yu[1], Ksenija Gasic[2], B. Todd Campbell[3], Katherine M. Evans[4], Cameron Peace[1] and Dorrie Main[1], (1)Washington State University, Pullman, WA, (2)Clemson University, Clemson, SC, (3)USDA-ARS, Florence, SC, (4)Washington State University, Wenatchee, WA

Breeding programs produce large amount of data that require efficient management systems to keep track of performance, pedigree, geographical and image-based data. With the development of DNA-based screening technologies, more breeding programs perform genotyping in addition to phenotyping for performance evaluation. The integration of breeding data with other genomic and genetic data is instrumental in the refinement of marker-assisted breeding tools, enhances genetic understanding of important crop traits and maximizes access and utility by crop breeders and allied scientists. We have previously developed a breeding database in the Genome Database for Rosaceae (GDR) and integrated with other genomic and genetic data. While it is developed using Chado and Drupal, it is not developed in Tripal. We report the progress on TripalBIMS, the Breeding Information Management System in Tripal which we have implemented in CottonGEN, the genetic, genomic and breeding database for cotton. We highlight future plans for the development of a comprehensive breeding management system in Tripal.

W969: Tripal Database Network and Initiatives

CAMbase: a Resource for CAM Plant Genomics

Robert Clifford Moseley[1], Xiaohan Yang[2] and Gerald A. Tuskan[2], (1)University of Tennessee-Knoxville, Knoxville, TN, (2)Oak Ridge National Laboratory, Oak Ridge, TN

Crassulacean acid metabolism (CAM) is a physiological adaptation of plants that live in environments under stress conditions, in particular, drought and heat stress. This specialized mode of photosynthesis facilitates increased water-use efficiency (WUE) in the plant, which has prompted the increase in research to define the genetic elements of CAM. To assist this effort, CAMbase is designed to be a versatile and comprehensible web portal focused on CAM plant species genomics data for integrative utilization of the increasingly large collection of CAM plant genomics resources. This web portal incorporates tools supporting the use of data produced by the CAM biodesign project, as well as publicly available data. It features an instance of WebApollo with Jbrowse for the *Kalanchoe laxiflora* genome to facilitate community annotation updating. Also, an instance of an eFP browser is available for comparative analysis of time-course gene expression in *K. laxiflora* leaves. As more genomics data are being produced in the CAM biodesign project, they will be incorporated for further development of the portal, such as allowing the visualization of gene expression between CAM plant species in the eFP browser all at once.

W970: Tripal Database Network and Initiatives

TreeGenes: Enabling Visualization and Analysis in Forest Tree Genomics

Emily Grau[1], Steven A Demurjian Jr[1], Hans Vasquez-Gross[2], Damian Gessler[3], Margaret Staton[4], Sook Jung[5], Alex Feltus[6], Dorrie Main[5], Stephen P. Ficklin[5], David Neale[7] and Jill Wegrzyn[8], (1)University of Connecticut, Storrs, CT, (2)University of California

Davis, Davis, CA, (3)University of Arizona, Tucson, AZ, (4)University of Tennessee, Knoxville, Clemson, SC, (5)Washington State University, Pullman, WA, (6)Genetics & Biochemistry, Clemson University, Clemson, SC, (7)Dept. Plant Sciences University of California Davis, Davis, CA, (8)Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT
The TreeGenes project is a collection of web-based resources designed to support the needs of the forest tree research community. These resources include an extensive curated database as well as custom tools to manage the flood of information resulting from high-throughput genomics projects. The goal is to capture raw data and metadata from sample collection through to the final analysis stage. TreeGenes allows users to track and manage next generation sequencing and genotyping studies, query and download datasets, and perform analysis through interfaces that connect directly to high performance computing resources.

The TreeGenes database holds genetic, phenotypic, and environmental data from over 1700 forest tree species. This data is delivered to the public through taxonomy views, literature searches, genetic maps (CMap), genome browsers (Gbrowse), and custom interfaces such as CartograTree. While much of the data is curated from external repositories, advanced integration of metadata supporting genotype:phenotype and genotype:environment association studies has been developed.

To enhance the utility of this resource, TreeGenes is currently is transitioning to Tripal. This transition will allow us to take advantage of features such as improved user management, data loaders, and standardized schemas provided through Drupal and Chado. Custom development will include a new Tripal module to streamline public submission of genetic association data, which will allow for greater flexibility and improved integration. We will also discuss the challenges and benefits of transitioning a mature database and website to Chado and Tripal.

W971: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

The Barley Genome

Nils Stein, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland, Germany
Barley is one of the most important cereal crop species. It is a close relative to wheat and rye. Its haploid genome size exceeds 5 Gigabases (Gbp), almost twice the size of any fully sequenced organism or crop species. The International Barley Sequencing Consortium (IBSC) started in 2006 a project to establish a map-based high quality reference sequence of barley. After reaching important intermediate results like (i) virtual gene order maps based on sorted chromosome survey sequencing, (ii) a whole genome shotgun draft sequence integrated to a genome-wide physical map and (iii) a draft sequence of the barley genome gene space based on pooled BAC sequencing, IBSC has come close to reach its 10 year milestone of a genome-wide BAC-by-BAC based genome sequence of barley. All seven barley chromosomes were sequenced based on short-reads sequencing-by-synthesis technology along their respective minimum tiling path of the physical map. Sequence data of more than 85,000 BAC clones was assembled and integrated with high density genetic marker information and 3D conformation capture sequencing data (HiC). As a result 4.6 Gbp of the 4.8 Gbp non-redundant sequences could be linearly ordered into seven pseudomolecules representing a first version (v1.0) reference sequence of the barley genome. A nano-channel electrophoresis based optical map of fluorescently labeled high molecular weight DNA was used to independently validate the physical integrity of the scaffolds of the pseudomolecules.

W972: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

Reference Sequence of the Genome of *Aegilops tauschii,* the Progenitor of the Wheat D Genome

Jan Dvorak[1], Olin D. Anderson[2], Jeffrey L. Bennetzen[3], Xiongtao Dai[4], Matthew W. Dawson[5], Karin R. Deal[1], Katrien M. Devos[6], Jaroslav Dolezel[7], Yong Q. Gu[8], Naxin Huo[9], Karl G. Kugler[10], Philippe Leroy[11], Yong Liang[12], Zhiyong Liu[12], Ming-Cheng Luo[1], Eric Lyons[13], Zhengqiang Ma[14], Long Mao[15], Klaus F.X. Mayer[16], W. Richard McCombie[17], Pat McGuire[1], Hans-Georg Mueller[4], Shuhong Ouyang[12], Geo Pertea[18], Daniela Puiu[18], Steven L. Salzberg[18], Carol Soderlund[19], Qixin Sun[20], Sven O. Twardziok[16], Hao Wang[21], Yi Wang[22], Zhenzhong Wang[12], Thomas Wicker[23], Lichan Xiao[12], Mengyuan Xiao[5], Frank M. You[24], Tingting Zhu[1] and Aleksey Zimin[25], (1)Department of Plant Sciences, University of California, Davis, Davis, CA, (2)Western Regional Research Center, Albany, CA, (3)Dept. of Genetics, University of Georgia, Athens, GA, (4)Dept. of Statistics, Davis, CA, (5)University of California, Davis, CA, (6)Institute of Plant Breeding, Genetics and Genomics (Dept. of Crop and Soil Sciences), and Dept. of Plant Biology, University of Georgia, Athens, GA, (7)Institute of Experimental Botany, Olomouc, Czech Republic, (8)USDA ARS, Western Regional Research Center, Albany, CA, (9)Dept. of Plant Sciences, Davis, CA, (10)Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany, (11)INRA UMR 1095 GDEC, Clermont-Ferrand, France, (12)China Agricultural University, Beijing, China, (13)School of Plant Sciences, iPlant Collaborative, Tucson, AZ, (14)Nanjing Agricultural University, Nanjing, China, (15)Chinese Academy of Agricultural Sciences, Beijing, China, (16)Helmholtz Center Munich - Plant Genome and Systems Biology, Neuherberg, Germany, (17)Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (18)Johns Hopkins University, School of Medicine, Baltimore, MD, (19)University of Arizona/Bio5 Institute, Tucson, AZ, (20)Northwest A&F University, Yangling, China, (21)Institute of Plant Breeding, Genetics and Genomics, and Dept. of Plant Biology, University of Georgia, Athens, GA, (22)Department of Plant Sciences, University of California, Davis, CA, (23)University of Zurich, Zurich, Switzerland, (24)Cereal Research Centre, Morden, MB, Canada, (25)University of Maryland, College Park, MD
The aim of this project (NSF IOS-1238231) is to generate a reference sequence for the genome of *Aegilops tauschii*, one of the diploid progenitors of hexaploid wheat. Because of the large size of the genome, we used the ordered-clone approach as the primary strategy to sequence the *Ae. tauschii* genome. We constructed a bacterial artificial chromosome (BAC)-based physical map of accession AL8/78, sequenced 42,025 BAC clones of a minimum tiling path across 3,578 BAC contigs and 2,000 singletons in eight-clone pools with the Illumina MiSeq platform, and assembled scaffolds with SOAPdenovo2. We also produced HiSeq 2500 whole-genome shotgun reads for accession AL8/78, which NRGene assembled with their DeNovoMAGIC™ software. We merged the two assemblies and generated assembly v2.0, for which the N50 of contigs and scaffolds were 93,210 and 2,884,388 bp, respectively. To validate the assembled scaffolds and assist with the construction of pseudomolecules, we constructed optical BioNano genome (BNG) maps of *Ae. tauschii* accessions AL8/78 and CIae23, merged them, aligned scaffolds against the BNG contigs, oriented and edited them, and produced seven pseudomolecules of a total length of 4,025,304,143 bp. Transposable element annotation revealed that 80% of the genome consists of transposable elements, which we classified into 3,326 families, of which 1,658 were newly described. Preliminary annotation suggested 37,000 high-confidence genes in the genome.

W973: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

Assessing Structural Variations Along the Bread Wheat Genome

Frederic Choulet[1], Hélène Rimbert[1], Ambre-Aurore Josselin[1], Emeric Dynomant[1], Valerie Barbe[2], Adriana Alberti[2], Karine Labadie[2], Sophie Mangenot[2], Arnaud Couloux[2], Arnaud Bellec[3], Jan Vrana[4], Marie Kubalakova[4], Jaroslav Dolezel[4], Hélène Bergès[3], Patrick Wincker[2] and Etienne Paux[1], (1)INRA GDEC, Clermont-Ferrand, France, (2)CEA - Genoscope, Evry, France, (3)INRA - CNRGV, Castanet Tolosan, France, (4)Institute of Experimental Botany, Olomouc, Czech Republic

Structural variations (SVs) represent a type of polymorphisms affecting genomes related to rearrangements such as insertions, deletions, duplications, inversions and translocations. In an attempt to decipher the extent of SVs in the hexaploid wheat genome, we have established strategies combining the whole-genome draft assembly of cv. Chinese Spring, the pseudomolecules of chromosomes 3B and 1B (under construction), and resequencing data from cultivated and wild wheat accessions, including whole genome shotgun from13 bread wheat accessions as well as whole-chromosome shotgun of flow-sorted chromosomes 3B from 44 hexaploid and tetraploid wheat accessions. Our results revealed high levels of SVs affecting genes and transposable elements (TEs) and reflecting the phylogeny. Local variations of the extent of SVs follow the structural and functional partitioning of chromosome 3B observed previously, with the chromosomal extremities having accumulated much more polymorphisms than the central part of the chromosome.

W974: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

Maize and Wheat Genomic Sequence Diversity Revealed using *de novo* Assembly and Unbiased Comparative Genomics

Gil Ronen, NRGENE Ltd., Ness-Ziona, Israel

A quality denovo assembly with long genomic sequences from short reads has long been considered an impossible challenge for large, repetitive genomes such as maize and wheat.

The challenge has been met and overcome by the DeNovoMAGIC2 genome assembly algorithm. Using high coverage (>180X) of short Illumina reads produced from a mix of library sizes, ranging from 450 – 10,000 bp, extremely long sequence scaffolds (L90>1 Million bp) with very low (<2%) unfilled gaps were produced for maize, diploid wheat and tetraploid wheat. These assemblies covered more than 90% of the genome sequences and could be produced with exceptional speed.

The unbiased assembly of multiple maize genomes enabled us to efficiently construct a "pangenome" database. Such a database facilitates the imputation of the full genome sequence of each individual maize line, as well as its haplotype map and polymorphism diversity, from genotyping data. As such, genomic-selection, GWAS and trait mapping are more simplified and accurate.

W975: Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics

Unlocking Bread Wheat Genome Diversity with New Sequencing and Assembly Approaches

Matt Clark, The Genome Analysis Centre, Norwich, United Kingdom of Great Britain and Northern Ireland

W976: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

Positional Cloning of Fhb1 Gene in Wheat

Nidhi Rawat[1], Michael Pumphrey[2], Sixin Liu[3], Xiaofei Zhang[4], Vijay K. Tiwari[1], Harold N. Trick[1], William W. Bockus[1], Eduard Akhunov[1], James A Anderson[4] and Bikram S. Gill[1], (1)Kansas State University, Manhattan, KS, (2)Washington State University, Pullman, WA, (3)USDA-ARS-NCCCWA, Kearnysville, WV, (4)University of Minnesota, St. Paul, MN

Fusarium Head Blight (FHB) also known as scab caused by ascomycetous fungus *Fusarium graminearum* [teleomorph *Gibberella zeae* (Schweinitz) Petch] is a destructive disease of wheat, barley, and other small grain crops. It leads to billions of dollars in crop losses worldwide. A quantitative trait locus (QTL) *Fhb1* located on chromosome 3B short arm of a few Chinese landraces is the most consistently reported and widely deployed source of resistance to scab. We report the positional cloning of *Fhb1* from Sumai 3, a Chinese cultivar resistant to scab. The QTL was previously mendelized in a Near Isogenic Line (NIL) set. A Sumai 3 BAC library with ~3x genome coverage was constructed, and markers derived from wheat reference landrace Chinese Spring spanning the QTL were used to screen the library. Four BACs forming the minimum tiling path were sequenced, and genes were annotated. Putative candidate genes were shortlisted and reverse genetics tools of Targeting Induced Local Lesions in Genome (TILLING) and RNAi-induced gene silencing were used to validate the gene. The results from the research will be presented.

W977: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

Rapid Cloning of Resistance Genes in Wheat

Burkhard Steuernagel[1], Sambasivam Periyannan[2], Inmaculada Hernández-Pinzón[3], Kamil Witek[3], Matthew N. Rouse[4], Guotai Yu[1], Asyraf Hatta[1], Michael Ayliffe[2], Evans Lagudah[2], Brande Wulff[1] and Jonathan Jones[3], (1)John Innes Centre, Norwich, United Kingdom, (2)CSIRO, Canberra, Australia, (3)The Sainsbury Laboratory, Norwich, United Kingdom, (4)USDA-ARS, University of Minnesota, St Paul, MN

To defend against disease, plants have a repertoire of resistance (R) proteins that can detect the presence of pathogen effector proteins. Genes encoding these proteins, which typically have a nucleotide binding (NB) site domain and leucine-rich repeats (LRR), are interesting targets for recruiting new resistance from wild relatives into crops. The traditional way to identify genes is by map-based cloning, which is limited by recombination in the target area. As a complementary approach, we have developed a method based on mutational genomics and R gene enrichment sequencing (RenSeq). An accession with a confirmed single resistance gene is subjected to an EMS mutant screen. Genomic DNA from mutants that show loss of resistance as well as from the wild-type is subsequently enriched for NB-LRR type genes using a custom biotinylated oligonucleotide bait library followed by identification of candidate genes through bioinformatics pipeline. As a proof of concept, we re-cloned the previously cloned gene *Sr33* from hexaploid wheat. From sequencing of six EMS mutants we unambiguously found a single candidate, which is identical to the real gene. We further cloned two stem rust resistance genes, *Sr22* and *Sr45*, that confer resistance to several

stem rust races including Ug99. In both cases, we identified a single, unambiguous candidate from six susceptible mutants, and for Sr22, we verified Sr22 function in transgenic wheat plants. Our method of using EMS mutants combined with RenSeq can efficiently clone NB-LRR type resistance genes. Due to the EMS based approach it relies on the presence of a single *R* gene in order to obtain susceptible mutants and it is limited to the gene space targeted by the bait library. The method, however, neither relies on positional fine mapping nor does it require a BAC library and therefore provides an excellent complementary approach to map-based cloning.

W978: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

A Toolbox to Clone QTLs for Host and Nonhost Resistances in Barley

Rients E. Niks, Wageningen University & Research Centre, 6700 AJ Wageningen, Netherlands

Cloning genes for nonhost resistance (NHR) and for quantitative host resistance (QHR) to specialized biotroph pathogens is a challenge. To date, only five large-effect resistance QTLs have been cloned for QHR, and none for NHR. We have developed the following tools to map-based clone genes for these two types of resistance in barley.

At seedling stage, barley (*Hordeum vulgare*) may show some degree of susceptibility to unadapted *Puccinia* rust fungi and *Blumeria graminis* powdery mildew fungi. We accumulated such susceptibility in the experimental line SusPtrit (Su), resulting in exceptional susceptibility to at least ten unadapted rust fungi. Standard cultivar Vada (Va) shows nonhost immunity to unadapted rust fungi and high partial resistance to the adapted *P. hordei*. In the Va/Su mapping population many QTLs for resistance to *Puccinia* rusts and *Blumeria*mildews were mapped. Phenotyping is based on inoculation in a settling tower, and counting infections as flecks or pustules (rusts) or micro-colonies (mildew). We find recombinants using EST based markers from a dense integrated linkage map of over 5000 SNP markers, in combination with high throughput genotyping by LightScanner® technology. We developed non-gridded BAC libraries for Va and for Su. Finally, we developed an experimental line, Golden SusPtrit, which combines the exceptional susceptibility of Su with the amenability for transformation of Golden Promise. This line can be transformed with candidate resistance genes to test their effect on level of infection.

Presently eleven NHR and QHR genes have been fine-mapped, of which four have been physically mapped to BAC clones.

W979: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

The Genetic Basis of Composite Spike Form in Barley and 'Miracle-Wheat'

Thorsten Schnurbusch, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland OT Gatersleben, Germany

Inflorescences of the tribe *Triticeae*, which includes wheat (*Triticum* sp. L.) and barley (*Hordeum vulgare* L.) are characterized by sessile spikelets directly borne on the main axis, thus forming a branchless spike. 'Compositum-Barley' and tetraploid 'Miracle-Wheat' (*T. turgidum* convar. *compositum* (L.f.) Filat.) display non-canonical spike-branching in which spikelets are replaced by lateral branch-like structures resembling small-sized secondary spikes. As a result of this branch formation 'Miracle-Wheat' produces significantly more grains per spike, leading to higher spike yield. In this study, we first isolated the gene underlying spike-branching in 'Compositum-Barley', i.e. *compositum 2* (*com2*). Moreover, we found that *COM2* is orthologous to the *branched head*^{t} (*bh*^{t}) locus regulating spike-branching in tetraploid 'Miracle-Wheat'. Both genes possess orthologs with similar functions in maize *BRANCHED SILKLESS 1* (*BD1*) and rice *FRIZZY PANICLE/BRANCHED FLORETLESS 1* (*FZP*/*BFL1*) encoding AP2/ERF transcription factors. Sequence analysis of the *bh*^{t} locus in a collection of mutant and wild type tetraploid wheat accessions revealed that a single amino acid substitution in the DNA-binding domain gave rise to the domestication of 'Miracle-Wheat'. mRNA *in situ* hybridization, microarray experiments, and independent qRT-PCR validation analyses revealed that the branch repression pathway in barley is governed through the spike architecture gene *Six-rowed spike 4* regulating *COM2* expression, while *HvIDS1* (barley ortholog of maize *INDETERMINATE SPIKELET 1*) is a putative down-stream target of *COM2*. These findings presented here provide new insights into the genetic basis of spike architecture in *Triticeae*, and have disclosed new targets for genetic manipulations aiming at boosting wheat's yield potential.

W980: Triticeae Genetics and Genomics, Session 2: Trait genetics and gene identification

The Non-brittle Rachis Genes of Triticeae

Takao Komatsuda, National Institute of Agrobiological Sciences, Tsukuba, Japan

W981: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement

Wheat Initiative

Helene Lucas, INRA Versailles, Versailles, France

W982: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement

Establishment of Exome-Captured TILLING Populations in Polyploidy Wheat and their use to Dissect Grain Size Components

Cristobal Uauy, John Innes Centre, Norwich, England

W983: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement

From Genes to Phenotypes in Wheat – Towards Breeder Friendly Markers

Curtis J Pozniak[1], Kirby Nilsen[1], P.R. MacLachlan[1], David Konkin[2], Andrew G. Sharpe[3], Krystalee Wiebe[1] and Sateesh Kagale[2], (1)University of Saskatchewan, Saskatoon, SK, Canada, (2)National Research Council Canada, Saskatoon, SK, Canada, (3)National Research Council Canada / Global Institute for Food Security (U of S), Saskatoon, SK, Canada

Modern selection technologies, mainly marker assisted selection (MAS), are used by our spring wheat (*Triticum aestivum* L.) and durum (*Triticum turgidum* var. *durum*) breeding programs to improve selection efficiency. However, most agronomically important traits are associated with quantitative trait loci (QTL) and only a few studies have defined genes associated with these in wheat. Next-generation sequencing

strategies and high density single nucleotide polymorphism (SNP) arrays have allowed improved resolution of QTL in well phenotyped bi-parental mapping populations and the recent availability of physical maps, and chromosomal/whole genome shot gun assemblies of the wheat genome promises the possibility to associate functional genes with these QTL. However, only a single wheat chromosome has been sequenced to reference quality (chromosome 3B) and current whole genome assemblies are still highly fragmented. This has slowed our map-based cloning efforts because much of the assembled scaffolds lack genetic position and order. Here we will report on our efforts to improve the current whole genome assembly of wheat that, in combination with transcriptome/exome sequencing and digital phenotyping, are aiding in marker development and postional cloning of a gene for insect resistance in both spring and durum wheat. Detailed genetic studies in both wheat species have identified two putative candidate genes, but it appears that SNPs within these do not capture all the phenotypic variation. Current data suggests that larger structural variants, coupled with gene silencing in some genetic backgrounds, may play an important role in trait expression.

W984: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
Predicting Grain Yield based on Predictor Traits from High-Throughput Phenotyping, Genome-Wide Markers, and Pedigree to Increase Rates of Genetic Gain per Unit Time and Cost
Jessica Rutkoski[1], Jesse Poland[2], Suchismita Mondal[3], Enrique Autrique[3], Lorena González Pérez[4], Jose Crossa[5], Matthew Reynolds[6] and Ravi P. Singh[5], (1)Cornell University & CIMMYT, Ithaca, NY, (2)Kansas State University, Manhattan, KS, (3)CIMMYT, Texcoco, Mexico, (4)CIMMYT, Ciudad Obregón, Mexico, (5)CIMMYT, Mexico DF, Mexico, (6)CIMMYT, Mexico D.F., Mexico
Genomic selection (GS) and high-throughput phenotyping (HTP) promise to revolutionize crop improvement. Physiological traits measured using field-based HTP platforms could be useful as yield predictors. We tested if using aerial measurements of canopy temperature (CT) and green and red normalized difference vegetation index (GNDVI, RNDVI) as predictor traits in pedigree and GS models could increase prediction accuracy for grain yield (GY). Predictor traits on training and test sets, and GY on the training set were modeled as multivariate and compared to univariate models with GY on the training set only. Predictor traits increased prediction accuracy for GY by 11% and 15% for genomic and pedigree models, respectively. Pedigree models with predictor traits were 17% more accurate than GS models without predictor traits. This suggests that predictor traits measured by HTP can be used to enhance GS or as an alternative to GS when pedigree relationships are known.

W985: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
TILLING Populations as Forward-Genetics Gene Machines
Silvio Salvi, DipSA - University of Bologna, Bologna, Italy

W986: Triticeae Genetics and Genomics, Session 3: Application of genomic resources to Triticeae improvement
WheatIS: A Genetics and Genomics Information System for the Wheat Research Community
Hadi Quesneville, INRA - URGI, Versailles, France
In 2011, the G20 ministers of agriculture mandated the Wheat Initiative to coordinate worldwide research efforts for wheat improvement. Under its umbrella, the WheatIS Expert Working Group was set up in 2013 to build an international wheat information system, called WheatIS (wheatis.org).
We present here the WheatIS conceived as distributed information system, acting as a hub for integrating wheat data produced and submitted to the public repositories. It relies on a network of 12 bioinformatics platforms working synergistically to provide an easy access to wheat data. These platforms, each considered as a WheatIS node, share their resources and propose several dedicated integrative databases, *e.g.* for genomic, genetic, and phenotype information, comparative genomics, and functional genomics. The hub, wheatis.org, provides centralized access to (i) the nodes and their resources, (ii) recommended data standards, (iii) a file repository to deposit and share the data among the scientific community, and (iv) a search in distributed databases for an easy data discovery.

W987: UCSC Genome Browser - a home for all organisms
UCSC Genome Browser - a Home for All Organisms
Robert Kuhn, U California Santa Cruz, Santa Cruz, CA
For 15 years the UCSC Genome Browser has been providing a visual display for genomic data from a large number of sequenced animals, now numbering 250 assemblies of more than 150 organisms. This workshop will demonstrate the Browser and a selected set of the most useful accompanying utilities: The Table Browser provides an intuitive interface on the massive underlying database; the Comparative Genomics datasets allow cross-species comparisons of mRNA and genomic alignments; custom tracks allow easily uploaded user data; and the Assembly Hubs mechanism allows anyone with sequence to display their organism on the UCSC platform

W988: Weedy and Invasive Plant Genomics
Ecological Genomics of Yellow Starthistle Invasions
Katrina Dlugosch, University of Arizona, Tucson, AZ
Yellow starthistle (Centaurea solstitialis) is a Eurasian species that has become highly invadive in the Americas. We have been reconstructing its routes of invasion, identifying major transitions in the evolution of increased plant size during invasion, and uncovering a potentially important role for structural evolution of the genome (genome size, content, and gene order) in the evolution of highly invasive populations, particularly in California.

W989: Weedy and Invasive Plant Genomics
Contrasting Recent, Continental Scale Invasions of the Americas by *Arabidopsis thaliana* and *Capsella bursa-pastoris*

Daniel Koenig[1], Christina Wesse[2], Moises Exposito-Alonso[3], Claude Becker[3], Barbara Neuffer[4] and Detlef Weigel[3], (1)University of California, Riverside, CA, (2)University of Osnabrück, Osnabrück, Germany, (3)Max Planck Institute for Developmental Biology, Tuebingen, Germany, (4)University of Osnabrü ck, Osnabrü ck, Germany
The migration of Europeans to the new world was accompanied by a rash of invasive flora that greatly disturbed the native ecological landscape. Colonizing plant populations frequently experienced strong founder effects, but they were nevertheless able to succeed in diverse habitats. I will discuss the demography and genetic basis for phenotypic variation in two such species, *A. thaliana* and *C. bursa-pastoris*. North American *A. thaliana* experienced an extreme founder effect with most individuals being derived from a single lineage. We have sequenced 76 genomes of extant individuals and 27 genomes of herbarium specimens from this lineage, together covering the time span from 1863 to 2006. Only three individuals show evidence of within-lineage recombination, greatly simplifying genetic analysis. We date the introduction to the mid-17th century, and we directly compare mutation rates throughout the genome. This comparison reveals that genetic drift predominates, but that purifying selection in this rapidly expanding population is nevertheless evident even over short historical times. Furthermore, a genome wide association study identifies mutations affecting traits likely to be under selection and important for adaptation in the wild. In contrast, genome-wide genetic analysis of 1000 *C. bursa-pastoris* collections suggests that adaptation to new environments was predominantly driven by standing variation. Climate stratified population structure is reiterated after colonization, and preliminary analysis suggests that the genetic basis of phenotypes relevant to climate adaptation can be mapped. Our analysis reveals alternative paradigms for success during recent invasions, and provides a glimpse into the genetic basis of adaptation on short timescales.

W990: Weedy and Invasive Plant Genomics

Sequencing the genespace of leafy spurge

David P. Horvath[1], James V. Anderson[1], Wun S. Chao[1], Munevver Dogramaci[1], Kevin Dorn[2] and Michael E. Foley[1], (1)USDA-ARS, Fargo, ND, (2)University of MN, St. Paul, MN
Leafy spurge is an invasive perennial weed that is a significant problem in range and recreational lands throughout much of the northern Great Plains states. An extensive sanger-sequenced EST database was developed that allowed the construction and use of a 23K element microarray. These microarrays were used for several significant studies on bud dormancy, invasiveness, and flowering of leafy spurge. More recently several RNAseq studies have investigated the impact of glyphosate on leafy spurge as well as additional studies on both season dormancy transitions and the transition from paradormancy to growth. These studies have yielded numerous clusters of coordinately regulated genes suitable for promoter analysis. However, leafy spurge lacked genomic sequences containing promoter elements. We have attempted to develop a gene space assembly of leafy spurge using a combination of transcriptome data along with 60X coverage of the leafy spurge genome with 100 and 150 base paired end read libraries. We utilized the Kmer selection program BBNorm to select fragments derived from low copy portions of the genome along with fragments that mapped to known transcribed genes. Initial analysis indicated these fragments contained both coding and non-coding regions of genes. These were assembled using the program Trinity, and Abyss. The Trinity assembly proved the best for with an N50 of 1353 and 102 million contigs.

W991: Weedy and Invasive Plant Genomics

A Comparative Sequence Analysis of Herbicide Target Genes across Dozens of Weedy Plant Species

Darci Giacomini, University of Illinois, Urbana, IL
As the price of sequencing has fallen drastically over the past decade, increasing amounts of herbicide target gene sequence data have become publically available to weed researchers. In an effort to consolidate this information, an online database is now accessible at http://weedscience.org, containing the cDNA sequences of 70+ weed species across seven herbicide target genes. This dataset provides a rich resource for weed scientists, both as a starting place for further target site resistance work in poorly studied species and as a data source for researchers interested in herbicide resistance evolution. Towards this second goal, gene and protein comparisons were carried out on each herbicide target gene, starting with the construction of multiple sequence alignments (MSAs) for all seven genes. These MSAs were superimposed on the protein crystal structure to generate conservation/variability scores at each residue and highlight regions that appear to be prone to mutation. Known herbicide resistance mutation sites were then identified in each MSA and likelihood scores for these mutations to appear for each weed species were calculated, allowing for a ranking of the species from highest to lowest risk of future herbicide resistance evolution. The utility of this analysis for weed managers will be discussed as well as other potential uses of the dataset.

W992: Weedy and Invasive Plant Genomics

The Genomic Landscape of the 297 Kilobase EPSPS Amplicon in Palmer Amaranth

William Molin, USDA-ARS, Stoneville, MS
Amaranthus palmeri, has evolved resistance to glyphosate by amplification of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene, which results in higher titers of EPSPS protein, the target site of glyphosate. A BAC library was generated using genomic DNA from a glyphosate resistant biotype. By sequencing overlapping BACs, a single consensus sequence of 297 kilobases was generated which included a single EPSPS gene of 8 exons and 7 introns in a 10,229 bp sequence. Approximately 140kb of genomic sequence was captured both 5' and 3' of the *EPSPS* locus. The genomic structure flanking the *EPSPS* locus was a complex configuration of tandem and inverted repeats. Also present were a ricesleeper homolog, a NAC domain containing protein, a heat shock cognate 70kD protein and a reverse transcriptase, all of which were transcribed. Whole genome shotgun sequencing (wgs) of two contrasting biotypes, sensitive and resistant, was performed on an Illumina platform and the reads were mapped to the reference amplicon, revealing significant differences in both repetitive and coding content between the biotypes. Alignment of wgs sequences from sensitive plants to the amplicon showed numerous gaps in the sequence, indicating that the amplified region was not contiguous in sensitive plants. The origin and ends of the amplicon and the mechanism of amplification remain to be identified. We propose that this amplicon of unprecedented length presents a unique adaption to herbicide stress and represents a mechanism for rapid evolution of herbicide resistance.

W993: Weedy and Invasive Plant Genomics

Population Genomics of Glyphosate-Resistant Palmer Amaranth (*Amaranthus palmeri*) Using Genotyping-by-Sequencing (GBS)

Anita Kuepper[1], **Harish Manmathan**[1], William McCloskey[2], Eric Patterson[1], Scott Nissen[1], Scott Haley[1] and Todd Gaines[1], (1)Colorado State University, Fort Collins, CO, (2)The University of Arizona, Tucson, AZ

Throughout the southeastern and southwestern United States, populations of Palmer amaranth (*Amaranthus palmeri*) have been identified with evolved resistance to the herbicide glyphosate. This project aims to determine the degree of genetic relatedness among a set of glyphosate-resistant and –susceptible lines by analyzing patterns of phylogeography and diversity on an intraspecific level. Seven different lines of Palmer amaranth from different geographic regions were tested against a glyphosate-resistant line from an Arizona locality for glyphosate resistance. The goal is to ascertain whether resistance evolved independently in the Arizona locality, or whether resistance spread from outside to the location. For example, the transportation of resistant seeds in harvesting equipment could be a source of gene flow via seed migration. The accumulation of shikimic acid via the shikimate assay and EPSPS copy number and were tested to confirm resistance. The susceptible lines showed an average of 41 mg/ml shikimic acid while the resistant lines showed an average of 0.1 mg/ml shikimic acid accumulation after exposure to a 500µm solution of glyphosate. Individuals from the Arizona glyphosate-resistant locality had increased copies of EPSPS in the range of 20 – 290-fold. This is the same mechanism previously identified in the Palmer amaranth lines from the southeastern US, therefore it is possible that resistance was introduced from elsewhere. DNA samples were collected for genotyping by sequencing (GBS) to perform single nucleotide polymorphism (SNP) calling, which will be used to determine the genetic structure of the different lines. Currently, neighbor joining trees and principle component analysis are being performed. This information about the evolution and migration of glyphosate resistance will be useful to design better strategies for herbicide resistance management.