Studying temperature and salinity stresses in the model brown alga Ectocarpus sp. by QTL mapping

Komlan Avia, Susana M. Coelho, Alexandre Cormier, Fiona Lercker, Stéphane Mauger, Gabriel J Montecinos, Sylvain Faugeron, Myriam Valero, J. Mark Cock, Pierre Boudry

5th International Conference on Quantitative Genetics

June 12-17, 2016

Madison, Wisconsin, USA

ICQG·5

Thank you to the following sponsors for their support of ICQG!

GOLD SPONSORS

AgReliant GENETICS

USDA
United States Department of Agriculture
National Institute of Food and Agriculture

SILVER SPONSORS

Aviagen

Cobb

DUPONT PIONEER

Genus

ABS

PIC

HENDRIX GENETICS

SEEDING THE FUTURE SINCE 1856

KWS

MONSANTO

syngenta

BRONZE SPONSORS

BAYER Bayer

DOW Dow AgroSciences

NEOGEN

WELCOME!

Welcome to the 5th International Conference on Quantitative Genetics (ICQG5) at the Monona Terrace Community and Convention Center in beautiful Madison, Wisconsin – USA!

For guaranteed admittance to all events, please wear your name badge at all times.

The keynote presentation will be held in Ballroom AB with the welcome reception following in the Grand Terrace. All other general sessions will be held in Exhibit Hall B. The exhibitors and poster sessions will be in Exhibit Hall A. The Wednesday happy hour will be held on the rooftop center circle (rain back-up is Ballroom AB and the Grand Terrace East.)

If you made a special dietary request, the ticket will be in the back of your name badge. Save this ticket and show it at all meal functions for specially requested items that meet your needs.

If you ordered a conference dinner ticket, the ticket is in the back of your name badge. Save this ticket for admittance to the conference dinner on Friday, June 17.

If you have any questions, during the conference, please ask at the ICQG registration desk.

Thank you and enjoy your visit to Madison!

CONTENTS

COMMITTEES

Local Organizing Committee:

Natalia de Leon (co-chair) – Department of Agronomy, Plant Breeding & Plant Genetics program

Guilherme Rosa (co-chair) – Departments of Animal Sciences and Biostatistics & Medical Informatics

Karl Broman – Departments of Biostatistics & Medical Informatics and Genetics

Jeff Endelman, Department of Horticulture

Corinne Engelman – Department of Population Health Sciences

Dan Gianola – Departments of Animal Sciences, Dairy Science and Biostatistics & Medical Informatics

Christina Kendziorski - Department of Biostatistics & Medical Informatics

David Page - Departments of Computer Sciences and Biostatistics & Medical Informatics

Bret Payseur – Department of Genetics

Kent Weigel – Department of Dairy Science

Brian Yandell – Departments of Statistics and Horticulture

ICQG International Steering Committee:

David Balding – University College of London, UK

Piter Bijma –Wageningen University, The Netherlands

Rachel Brem – Buck Institute, Novato, USA

Ed Buckler – USDA/Cornell University, Ithaca, USA

Andy Clark – Cornell University, Ithaca, USA

Mark J. Daly – Harvard University, Boston, USA

Rebecca Doerge – Purdue University, West Lafayette, USA

Jonathan Flint – University of Oxford, UK

Greg Gibson – Georgia Tech University, USA

Bill Hill - University of Edinburgh, Scotland

Ary Hoffmann - The University of Melbourne, Australia - Frederic Hospital – INRA, France

Trudy Mackay – North Carolina State University, Raleigh, USA

Albrecht Melchinger – University of Hohenheim, Germany

Bill Muir – Purdue University, West Lafayette, USA

Patrick Phillips – University of Oregon, Eugene, USA

Daniel Pomp – University of North Carolina, Chapel Hill, USA

Pak Chung Sham - The University of Hong Kong, Hong Kong

Fred VanEeuwijk - Wageningen University, The Netherlands

Peter Visscher – The University of Queensland, Australia

Bruce Walsh – University of Arizona, Tucson, USA

Bruce Weir – University of Washington, Seattle, USA

Qifa Zhang - Huazhong Agriculture University, China

SPONSOR INFORMATION

Thank you to the ICQG 5 sponsors!

GOLD

AgReliant Genetics

Headquartered in Westfield, Ind., AgReliant Genetics is an innovative seed company committed to delivering high quality seed, providing exceptional service and creating consistent customer value. Created in 2000 as a joint venture between the international seed groups KWS and Limagrain, AgReliant Genetics is ranked as one of the largest field seed companies in North America. AgReliant Genetics markets corn, soybean and alfalfa seed through eight brands: AgriGold®, Eureka Seeds®, Great Lakes Hybrids®, Golden Acres® Genetics, LG Seeds®, PRIDE Seeds®, Producers Hybrids® and Wensman Seed®.

More information: http://agreliantgenetics.com/

United States Department of Agriculture – National Institute of Food and Agriculture (USDA NIFA)

NIFA supports research, educational, and extension efforts in a wide range of scientific fields related to agricultural and behavioral sciences. In all of these areas, you will find NIFA working in pursuit of our vision. To address contemporary agricultural challenges, we seek to catalyze transformative discoveries and enhance education and engagement.

NIFA programs are the mechanism by which we pursue our mission to invest in and advance agricultural research, education, and extension to solve societal challenges. Some programs administer funding and offer leadership while others focus on national leadership and collaboration.

More information: https://nifa.usda.gov/

SILVER

Aviagen

Aviagen Broiler Breeders supplies day-old grandparent and parent stock chicks to customers in 130 countries worldwide under the Arbor Acres®, Indian River®, and Ross® brand names. These brands are among the most recognized and respected names in the industry and each has a proven record of success in addition to a large and loyal global customer base.

Aviagen also offers specialty breeding stock aimed to give customers flexibility in their product choice and meet specific market requirements. The Rowan Range® brand of specialty birds, with variable coloring depending on female choice, *currently only available in Europe*, meets the needs of selected niche or emerging markets, including the slower-growing, free-range, and organic segments. The Yield Plus male is the first available product in the Specialty Male portfolio from Aviagen and is intended to produce an end product with significantly more breast meat yield.

Aviagen's successful and well-established genetic selection program consistently promotes continuous improvements in robustness and overall health while providing the birds with the highest quality care and welfare standards.

More information: http://en.aviagen.com/

Cobb-Vantress Inc.

Cobb-Vantress, Inc. is a poultry research and development company engaged in the production improvement and sale of broiler breeding stock. Cobb is the world's oldest pedigree broiler breeding company. Founded in 1916, Cobb has grown into one of the world's leading suppliers of broiler breeding stock with distribution into over 100 countries. Cobb has contributed to the dynamic efficiency and growth of an industry that has transformed chicken into an economically affordable healthy protein source for many of the world's almost 7 billion people.

More information: http://www.cobb-vantress.com/

DuPont Pioneer

DuPont Pioneer is the world's leading developer and supplier of advanced plant genetics providing high-quality seeds to farmers around the world.

With business operations in more than 90 countries, DuPont Pioneer develops and distributes high-quality corn, soybeans, sorghum, sunflower, alfalfa, canola, wheat, rice, cotton, pearl millet and mustard seed, as well as forage additives, and a variety of services and expertise.

A demonstrated partner to farmers, Pioneer is committed to increasing food production with high quality Pioneer® brand products and agronomic knowledge that maximizes agricultural productivity. By combining conventional and new technologies, DuPont Pioneer is delivering solutions to help meet the needs of a growing population whose demand for agricultural seeds and products continues to increase.

More information: http://www.dupont.com/

Genus

Pioneering Animal Genetic Improvement: Genus is a world-leading animal genetics company. We provide farmers with superior genetics that enable them to more efficiently produce higher-quality animal protein, in the form of meat and milk. Genus is a market leader in porcine, dairy and beef genetics and is uniquely positioned as a global player, with a dedicated, multi-species research and development ('R&D') function and an international distribution network.

More information: http://www.genusplc.com/

Hendrix Genetics B.V.

Hendrix Genetics is a leading multi-species breeding company with primary activities in layer, turkey, pig, aquaculture and traditional poultry breeding.

Hendrix Genetics is dedicated to generating solutions for the animal protein sector that meet the challenges of food production. Backed by a strong portfolio of leading brands, Hendrix Genetics provides expertise and resources to producers globally, with operations and joint ventures in 24 countries, and more than 2,500 employees worldwide.

More information: http://www.hendrix-genetics.com/

KWS

KWS is one of the world's leading plant breeding companies

We are contributing to overcoming the challenges which arise from a growing world population and the rising demand for food which it entails by improving the performance of our plant varieties and continuously refining the quality of our seed.

KWS' core competence is crop breeding. We use state-of-the-art plant breeding methods and technologies to continuously improve yields and resistances to diseases, pests and abiotic stress.

Its annual expenses for research and development amount to 17 percent of the entire net sales. The expenditure on research and breeding has increased continuously over the past ten years.

More information: http://www.kws.com/

Monsanto

Monsanto is a sustainable agriculture company. We deliver agricultural products that support farmers all around the world.

We are focused on empowering farmers—large and small—to produce more from their land while conserving more of our world's natural resources such as water and energy. We do this with our leading seed brands in crops like corn, cotton, oilseeds and fruits and vegetables. We also produce leading in-the-seed trait technologies for farmers, which are aimed at protecting their yield, supporting their on-farm efficiency and reducing their on-farm costs.

We strive to make our products available to farmers throughout the world by broadly licensing our seed and trait technologies to other companies. In addition to our seeds and traits business, we also manufacture Roundup® and other herbicides used by farmers, consumers and lawn-and-garden professionals.

More information: http://www.monsanto.com/

Syngenta

Syngenta is a leading agriculture company helping to improve global food security by enabling millions of farmers to make better use of available resources. Through world class science and innovative crop solutions, our 28,000 people in over 90 countries are working to transform how crops are grown. We are committed to rescuing land from degradation, enhancing biodiversity and revitalizing rural communities.

More information: http://www.syngenta.com/

BRONZE

Bayer

Beautiful fields of healthy, high-yielding crops. Abundant harvests of golden grains, white cotton and succulent produce – plentiful enough to nourish and clothe the world. Healthy environments in which we safely and comfortably live, work and play. At Bayer CropScience, these ideals drive us every day. Our singular purpose is to propel farming's future, harnessing cutting-edge agricultural and environmental innovations to deliver on Bayer's mission: Science For A Better Life.

More information: https://www.cropscience.bayer.us/

Dow AgroSciences

We are passionate about developing sustainable chemical and biotechnology solutions that minimize risk and help growers get the most out of every acre or hectare of land. We focus on increasing crop productivity through improved plant genetics and pest management solutions.

"We are the Science serving the needs of a growing world".

Dow AgroSciences is a wholly owned subsidiary of The Dow Chemical Company

More information: http://www.dowagro.com

Neogen's Geneseek

Neogen's GeneSeek subsidiary is the leading agricultural genomics service lab in the world and has been providing comprehensive genomic and diagnostic solutions since 1998. GeneSeek extends its expertise around the globe offering DNA testing services for animal and plant breeding and genetics, offering high throughput solutions at low cost.

More information: http://www.neogen.com/Genomics/

EXHIBITORS

Cobb-Vantress Inc.	**More information:** http://www.cobb-vantress.com/
Neogen's Geneseek	**More information:** http://www.neogen.com/Genomics/
Novogene Corporation	**More information:** http://www.novogene.com

SUNDAY, JUNE 12
ICQG OPENING SESSION

2:00 – 8:00 PM Registration Open (Monona Terrace Community and Convention Center Grand Terrace)

6:00 – 8:00 PM Welcome

6:00 – 6:20 PM Welcome Remarks:
Dr. Marsha Mailick (Vice Chancellor for Research and Graduate Education, UW-Madison)
Drs. Natalia de Leon and Guilherme Rosa, UW-Madison, ICQG5 co-chairs

6:20 - 7:00 PM Keynote Presentation: Ockham's razor- When is the simpler theory better?

Dr. Elliott Sober[1]
[1]*University of Wisconsin-Madison*

Ockham's razor, the principle of parsimony, says that a theory that postulates fewer entities, causes, or processes is "better" than a theory that postulates more, so longer as the simpler theory is compatible with what we observe. But what does "better" mean? It is obvious that simpler theories are easier to remember, manipulate, and test. The hard problem is to say why the fact that one theory is simpler than another is relevant to deciding what the world is like. In this lecture I'll describe two or three "parsimony paradigms" within which this hard problem can be solved.

7:00 - 8:00 PM Reception

MONDAY, JUNE 13
8:30 – 11:50 AM SESSION I:
QUANTITATIVE GENETICS IN THE GENOMICS ERA

8:30 – 8:40 AM Chair Remarks: Dr. Natalia de Leon, UW-Madison

8:40 – 9:20 AM Epic fails in quantitative genetics

Dr. Bruce Walsh[1]
[1]*University of Arizona*

The health and vibrancy of any field of scientific endeavor is better gauged by its failures and missteps than by it successes. Apparent paradoxes and unresolved issues imply a healthy and growing field, and by this standard, quantitative genetics is indeed doing well. I'll review a number of unresolved, and often contradictory, issues, as well as a number of false starts and directions where time and resources would have been better spent elsewhere (the WOMBAT -– waste of money, brains, and talent -– standard). Hopefully, this will set the stage for a robust debate throughout the congress on what quantitative genetics has done right, where it has come up short, and how best to move forward.

9:20– 10:00 AM Charting the genotype-phenotype map: Lessons from Drosophila

Dr. Trudy Mackay[1]
[1]*North Carolina State University*

Quantitative traits are affected by multiple interacting loci with individually small and environmentally sensitive effects. Knowledge of the detailed genetic architecture of quantitative traits is important from the perspectives of evolutionary biology, human health and plant and animal breeding. Understanding the genetic architecture of quantitative traits involves identifying the genes and naturally occurring genetic polymorphisms affecting variation in these traits and determining their frequencies and additive, dominance, epistatic and pleiotropic effects, in a range of relevant environments. The model organism Drosophila melanogaster brings an impressive toolkit to the challenge of genetically dissecting quantitative traits. Drosophila organismal and gene expression quantitative traits have complex genetic architectures, with large numbers of pleiotropic genes that have sex-, environment- and genotype-specific effects. These observations offer valuable lessons for understanding the genetic basis of variation for quantitative traits in other organisms, including humans.

10:00 – 10:30 AM Coffee break (30 min)

10:30 – 11:10 AM Inference and prediction in modern animal quantitative genetics

Dr. Daniel Gianola[1]
[1]*Departments of Animal Sciences, Biostatistics & Medical Informatics, and Dairy Science, University of Wisconsin-Madison*

Arguably, a history of statistical methods in animal quantitative genetics could have three tomes: I: “Fisher-Henderson”; II: “Henderson-Bayes”; III: “Bayesian genomics”. Volume I would narrate how a mean evolved to a vector (Xβ), how additive genetic variance morphed into a variance-covariance matrix, and how methods for likelihood-identified estimands were developed for inference. II would describe how animal breeders “estimated” breeding values for individuals without phenotypes using conditional distributions induced by genealogies, e.g., with BLUP. It would also tell us how neo-Bayesian ideas entered animal breeding, with MCMC as epitome: any model is implementable by sampling. III would inform about the emergence of massive amounts of DNA data, and on how Bayesian methods could be used for genomic selection and for unraveling “genetic architecture”. The preceding quest aims to relate a huge number of molecular variants to unknown quantitative trait loci using statistical models with more unknowns (p) than data points (n). III would also describe how cross-validation (CV) calibrates prediction machines, including non-parametric models, against observable outcomes. Bayesian methods address the curse of dimensionality elegantly, but priors can have a say in inference in uncertain manners. Volume III developing stories that merit deeper consideration are discussed: 1) Genome-based models for complex traits have patent specification errors, so what is being estimated? 2) What does a prior distribution over substitution effects mean? 3) What is the influence of a prior on inferences, and where is the information on n-p unknowns coming from? 4) Is a posterior distribution with millions of dimensions well estimated from a few thousand data points? 5) How much sampling is needed to impede Monte Carlo error overwhelm signal? 4) How much CV is needed if predictive distributions are to be estimated in a credible manner?

11:10 – 11:50 PM Measuring population structure and relatedness with genomic data

Dr. Bruce Weir[1]
[1]*University of Washington*

Whole-genome variant data are allowing detailed characterization of associations of alleles within individuals (inbreeding), between individuals within populations (relatedness) and between populations (F-statistics). At each level there is an implication that association is measured relative to that in some reference set. With this perspective, we can re-visit the estimation of F-statistics and express estimates as functions of allelic matching within and between populations. We find it helpful to provide population-specific estimates as matching within a population relative to the unweighted average matching between all pairs of populations in a study. For equal sample sizes, or populations in a star phylogeny, the estimates reduce to the standard ones. These measures can be used to detect the effects of natural selection or to identify regions associated with disease. Using very rare variants, including private variants, also allows statements to be made about the timing of past evolutionary events. The machinery set up for quantifying population structure carries over immediately to the estimation of individual inbreeding coefficients and pairwise kinship coefficients. These new estimates avoid the biases caused by using sample allele frequencies in place of population frequencies in many standard estimates. This unified approach to relatedness at the population and individual level is illustrated with data from the 1000 Genomes project.

11:50 – 1:20 PM Lunch

1:20 – 4:30 PM SESSION II:
ADVANCES IN MODELING AND COMPUTATIONAL TOOLS

1:20 – 1:30 PM Chair Remarks: Dr. Rebecca Doerge, Purdue University, USA

1:30 – 2:10 PM Post-translational mechanisms buffer protein abundance against transcriptional variation

Steven C. Munger[1,2], Joel M. Chick[1,3], Petr Simecek[2], Edward L. Huttlin[3], Kwangbom Choi[2], Daniel M. Gatti[2], Narayanan Raghupathy[2], Steven P. Gygi[3,4], Gary A. Churchill[2,4]
[1]*Co-first author,* [2]*The Jackson Laboratory, Bar Harbor, ME 04609, USA,* [3]*Harvard Medical School, Cambridge, MA, USA,* [4]*Co-senior author*

Recent studies have reported low correlation among transcript and protein levels in mouse tissues and human lymphoblastoid cell populations (Wu L. et al. Nature 2013, Wu Y. et al. Cell 2014, Battle et al. Science 2015). Here we describe the largest combined RNA-seq and shotgun proteomics study to date, in liver samples from genetically heterogeneous Diversity Outbred (DO) mice. We observe two classes of gene regulation that act on protein levels. Most protein abundance QTL (pQTL) map locally to the gene region and act in cis to affect both message and protein, resulting in high concordance between transcript and protein levels. In stark contrast, proteins with distant pQTL appear largely uncoupled from transcript abundance. We apply a novel conditioning approach to identify causal regulatory proteins and transcripts underlying distant pQTL. We discover that stoichiometric buffering of protein complexes and metabolic pathways is the predominant trans-acting post-translational mechanism that buffers protein levels against cis genetic variation in the protein-coding gene.

2:10 – 2:50 PM Next generation sequencing, transcriptomics, and phylogenetics: a case study with the Caryophyllales

Dr. Stephen Smith[1]
[1]*University of Michigan*

Next generation sequencing offers scientists the ability to explore the intersection of macroevolution and genomic patterns on a scale never before possible. However, many studies have primarily focused on the better resolution of phylogenetic relationships or on better understanding a small number of gene families. Here, I explore other analyses that these data can allow. Specifically, using a data set of over 200 transcriptomes and three genomes across the plant group the Caryophyllales, I explore genome and gene duplication events, discordant relationships, and the intersection of these with diversification. In order to examine these patterns, many new methods and techniques were created. I will also discuss how the creation of new methods is required to better utilize these new data resources.

2:50 – 3:05 PM Selected Poster Presentation: Genomic data integration for GWAS and eQTL analysis

Constanza Rojo[1], Qi Zhang[2], Sunduz Keles[1]
[1]*University of Wisconsin-Madison,* [2]*University of Nebraska Lincoln*

Genome wide association studies (GWAS) are widely used to elucidate genetic variation associated with traits. When combined with genome-wide expression analysis (eQTL), these studies have the potential to identify genetic variants that modulate disease or trait phenotypes through their impact on expression of genes. Challenging aspects of these studies are: high dimensionality of the variant space and that most candidate genetic variants are located either in intronic or intergenic regions making their interpretation challenging. We develop a novel mediation analysis framework to integrate eQTL and GWAS studies with the effective utilization of publicly available annotation data. This approach extends the scope of standard mediation analysis framework adapted by statistical genetics from accommodating 10s of genetic variants to 100s to 1000s of variants. The integrative framework incorporates the regulatory information encoded in annotation data to elucidate how top-ranking association SNPs might be modulating gene expression. We apply our method to Framingham Heart Study data of type 2 diabetes and identify expression modulating non-genic genetic variants with significant impact on diabetes.

3:05 – 3:35 PM Coffee break (30 min)

3:35 – 3:50 PM Selected Poster Presentation: Exploring properties of genome wide association analyses based on broadly different prior specifications

Chunyu Chen[1], Juan P. Steibel[1], Robert J. Tempelman[1]
[1]*Michigan State University*

Currently popular methods used for genome-wide association (GWA) studies are based on simultaneously fitting all genetic markers in an attempt to account for population structure. A currently popular strategy (EMMAX) infers upon association for the marker of interest by treating its effect as fixed (i.e., flat prior) with all other markers treated as Gaussian random effects. We conjecture that there may be merit to specifying the markers of inferential interest as also sharing the same prior distribution as all other markers in a joint analysis, whether that distribution is Gaussian (rr-BLUP), a heavier-tailed Student t (BayesA) or variable selection specifications based on a mixture of two Gaussian distributions (SSVS). We simulated ten replicated datasets for each of 9 = 3x3 populations defined by three different literature-based QTL distributions (gamma shapes of 0.18, 1.48, or 3) and three different total numbers of QTL (30, 90, or 300) using genotypes derived from

24,661 SNP markers on 928 individuals based on a Duroc-Pietrain F2 cross at Michigan State University. We quantified receiver operating characteristic (ROC) properties for each model by computing the area under the ROC curve (AUC) up until a false positive rate of 0.05 using 10 SNP or 1MB window based inference. We found that SSVS lead to the best (P<0.0001) AUC in all cases except for the most oligogenic and skewed QTL effects. EMMAX was generally the most inferior having an average AUC 60% that of SSVS. We also evaluated marginal a posteriori (MAP) approaches to BayesA and SSVS as an alternative to using MCMC; however, these approaches did not appear to be as promising due to their sensitivity to starting values. It appears then that there are considerable advantages to use variable selection specifications in GWA studies.

3:50 – 4:30 PM Multi-omic prediction of breast cancer progression using Bayesian generalized additive models

Ana I Vazquez[1], Agustin Gonzalez[1], Yogasudha Veturi[2], Gustavo de los Campos[1,3]
[1]: *Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI.*
[2]: *Department of Biostatisitcs, University of Alabama at Birmingham, Birmingham, AL.*
[3]: *Department of Statistics and Probability, Michigan State University, East Lansing, MI.*

Multi-omic information can be used to uncover disease processes and to develop statistical models for risk assessments. Here we describe a statistical framework for integrating multi-layer high-dimensional omics with clinical covariates for prediction of disease risk. We use data from The Cancer Genome Atlas and from the Molecular Taxonomy of Breast Cancer Consortium to develop models for prediction of breast cancer survival. Our results show that whole-genome gene expression (WGGE) profiles, and whole-genome methylation are more predictive of survival than any of the commonly used clinical covariates, including for example cancer subtype and stage. Copy Number Variants were also predictive of survival but their predictive power was lower than that of WGGE in both data sets. We also demonstrate how prediction of survival can be improved by combining clinical information with multi-omic data.

4:30 – 6:30 PM Poster Session and Social Hour

TUESDAY, JUNE 14

8:30 AM – 12:05 PM SESSION III:

THE GENETIC ARCHITECTURE OF QUANTITATIVE TRAITS

8:30 – 8:40 AM Chair Remarks: Dr. Henner Simianer, Georg-August University Goettingen, Germany

8:40 – 9:20 AM Genetic effects on molecular phenotypes reveal disease biology

Dr. Emmanouil Dermitzakis[1]
[1]*University of Geneva Medical School and Swiss Institute of Bioinformatics*

Molecular phenotypes inform us about genetic and environmental effects on cellular and tissue state. The elucidation of the genetic basis of gene expression and other cellular phenotypes is highly informative for the impact of genetic variants in the cell and the subsequent consequences in the organism. In this talk I will discuss recent advances in key areas of the analysis of the genomics of gene expression and cellular phenotypes in human populations and multiple tissues and how this assists in the interpretation of regulatory networks and human disease variants. I will also discuss how the recent advances in next generation sequencing and functional genomics are informing us about the impact of regulatory variation in cancer.

9:20 – 10:00 AM Linking genome wide association studies (GWAS) and genomic selection (GS) to better utilize natural variation in rice

Prof. Susan McCouch[1], Dr. Mark Wright[1], Dr. Jen Spindel[1], Chih-Wei Tung[1], Dr. Lyza Maron[1], Dr. Sam Crowell[1], Genevieve DeClerck[1], Pavel Korniliev[1], Anthony Greenberg[1], Francisco Agosto Perez[1], Dr. Namrata Singh[1], Dr. Randy Clark[2], Dr. Leon Kochian[2], Dr. Anna McClung[3], Dr. Georgia Eizenga[3], Jason Mezey[1]
[1]Cornell University, [2]USDA-ARS, Robert W. Holley Center, [3]USDA-ARS, Dale Bumpers National Rice Research Center

Understanding the relationship between genotypic and phenotypic variation lies at the heart of the study of genetics and is also critically important to applications in plant breeding. Here we present a genome-wide association study (GWAS) based on genotyping a rice diversity panel with a high-density SNP array and systematically phenotyping the panel for a range of agronomic, physiological and morphological traits. Seeds conserved in global germplasm repositories, including wild rice ancestors, ancient landrace varieties and modern elite cultivars provide insight into a wide range of natural genetic variation that remains underutilized in modern rice genetics and breeding. Using high throughput sequencing and genotyping technologies, we examine genome-wide patterns of variation and document deep sub-population structure within Oryza sativa. We use GWAS to identify common variants influencing complex traits and demonstrate heterogeneity of genetic architecture across subpopulations and environments. We use genomic selection (GS) in combination with de novo GWAS to predict breeding values and improve the rate of genetic gain. This work establishes an open-source translational research platform for genome-wide association studies and genomic selection in rice that directly links molecular variation in genes and metabolic/regulatory pathways with the germplasm resources needed to accelerate varietal development and crop improvement for diverse environments.

10:00 – 10:30 AM Coffee break (30 min)

10:30 – 10:45 AM Selected Poster Presentation: Characterizing the allele-frequency spectrum of causal variants in yeast

Dr. Joshua S. Bloom[1,2], Laura Day[1,2], Holly Oates-Barker[1,2], Dr. Sebastian Treusch[3], Dr. Leonid Kruglyak[1,2]
[1]UCLA, [2]Howard Hughes Medical Institute, [3]Twist Bioscience

For many heritable traits, including common diseases in humans, the relative contribution of variants in different allele frequency classes to trait variation remains a subject of debate. We examined the genetic architectures of 41 quantitative traits in a whole-genome-sequenced panel of 13,500 yeast strains. The panel comprises 16 large biparental crosses among 16 diverse yeast strains,employing a round-robin design. We detected between 43 and 155 QTL per trait and explained most of the additive trait variance with detected QTL. We observe a large range of architectures across the different traits, with the rarest variants explaining an average of 60% of additive variance per trait (but ranging from 10% to 97% per trait). For many traits, a large fraction of trait variation is explained by many low-frequency variants with small effects. In addition, we observed widespread allelic heterogeneity within the mapped loci for some traits. Our highly powered design provides the clearest picture yet of the architecture of complex traits and the allele-frequency spectrum of causal variants, and informs the design and interpretation of future mapping studies.

10:45 – 11:25 AM Whole Genome sequencing: From the identification to the prediction of quantitative trait nucleotides in pigs

Dr. Martien Groenen[1]
[1]*Wageningen University*

Genome wide association studies and selective sweep analysis in combination with whole genome sequencing, has enabled the identification of many sequence variants underlying important traits under selection in livestock species. In particular in pigs, the 19th century introgression of divergent Asian haplotypes into European commercial breeds, has proven to be a rich source of information about the genetic variants under selection in these breeds. By using an approach we refer to as "introgession mapping" we were able to identify many regions in the genome of the pig that most likely contain genetic variants involved in fatness and fertility, two traits that have been under strong selection in pigs. However, these approaches only enable the identification of a small fraction of the genetic variants underlying these traits. To fully utilise the wealth of information of the increasing amount of genome sequence information and to move towards biology-driven genomic prediction we need to be able to predict the potential functionality of every nucleotide variant in the genome. The recent establishment of the Functional Annotation of Animal Genomes (FAANG) consortium to characterize the non-coding sequences and epigenetic modifications in the genome are a first step towards this goal. We recently have started to apply machine learning approaches to predict the likelihood of a given variant to be deleterious or pathogenic, by integrating genome annotation, evolutionary conservation and functional genomics data, and transferring knowledge obtained from other species (human).

11:25 – 12:05 PM Breeding 4.0? Sorting through the adaptive and deleterious variants across the maize genome

Dr. Edward Buckler[1,2], Alberto Romero[2], Karl Kremling[2], Peter Bradbury[1,2], Sarah Hearne[3], Eli Rodgers-melnick[2], CIMMYT SEED Project, US Maize Diversity Project
[1]*USDA-ARS,* [2]*Institute for Genomic Diversity, Cornell University,* [3]*International Maize and Wheat Improvement Center (CIMMYT)*

Over much of the last two decades, amazing strides have been made both in molecular genetics and quantitative genetics. The genomic selection has already provide a powerful sets of tools that can advance breeding efficiently. However, the next decade is providing the tools for rapid and powerful genome editing - while this be the forth revolution of plant breeding? This will shift the plant genetics and breeding community from finding useful haplotypes to finding the credible sets of causative variants sets that can be edited. Using the tremendous diversity and rapid LD decay of maize and related grasses, a four pronged approach to dissecting variants is making substantial progress. (1) High throughput expression profiling is identifying deleterious variants and controllers of expression that appears to have direct implications for yield. (2) Surveys of thousands of maize landraces are pinpointing the adaptive alleles diverse environment. (3) Chromatin structure analysis is highlighting that only 2% of the genome is key for regulation. (4) Comparative genomics across the monocots are recommending some of the deleterious variants likely responsible for heterosis. The challenge for the next several years will be defining how credible are these variants sets, how do we refine them, and how accurate do they need to be in order to allow genome editing to be faster than genomic selection?

12:05 – 1:20 PM Lunch

1:20 – 4:30 PM SESSION IV:
NETWORKS AND CAUSALITY IN GENETICS

1:20 – 1:30 PM Chair Remarks: Dr. Brian Yandell, UW-Madison, USA

1:30 – 2:10 PM Graphical modeling in genetics

Dr. Peter Spirtes[1]
[1]*Carnegie Mellon*

The past 30 years has seen an explosion of research in the theory and application of graphical models to a variety of scientific domains. I will give an introduction to graphical models, including their representation, the assumptions underlying them, their use in both causal and predictive models, and current search algorithms that have been devised for constructing graphical models from data and background knowledge. I will use some case studies to describe special challenges that genetic data presents for the application of graphical models, how some of those challenges have been successfully dealt with, and future prospects.

2:10 – 2:50 PM Bayesian networks, MAGIC populations and multiple trait prediction

Dr. Marco Scutari[1]
[1]*University of Oxford*

Multiparent advanced generation inter-cross (MAGIC) populations represent a key resource in dissecting the genetic architecture and the interplay of phenotypic traits because of their high genetic recombination, diversity and large sample size. Bayesian networks provide a convenient and interpretable framework for simultaneously modelling those traits, taking into account the fact that they are most likely associated due to pleiotropy and shared biological basis.

In this talk we will discuss how to learn such models and we will explore their predictive accuracy using a winter wheat (Triticum aestivum L.) MAGIC population from NIAB and an indica rice (Oryza sativa) MAGIC population from IRRI. We will also provide examples of their ability of addressing confounding between traits and to perform inference on both traits and markers. Overall, we argue that Bayesian networks provide an intuitive model for both posterior and causal reasoning in genetics while at the same time providing competitive predictive accuracy.

2:50 – 3:05 PM Selected Poster Presentation: Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations

Mrs. Tessa Brinker[1], Dr. Esther D Ellen[1], Dr. Roel F. Veerkamp[1], Dr. Piter Bijma[1]
[1]*Wageningen University and Research – Animal Breeding and Genomics Centre*

Minimizing bird losses is important in the poultry industry. Selection against mortality is challenging because heritability is low, censoring is high, and survival depends on social interactions. With cannibalism, mortality depends on Direct Genetic Effects (DGE), i.e. effects of an individual's genotype on its own phenotype and Indirect Genetic Effects (IGE), i.e. effects of an individual's genes on trait values of its social partners. Studies using DGE-IGE models focussed on analysing survival time where censored records were considered as exact and it was assumed that IGE were continuously expressed, even after death. However, dead animals no longer express IGE, and IGE should therefore be time-dependent in the model. Neglecting censoring and timing of IGE expression may reduce accuracy of estimated breeding values (EBV). Thus, our aim was to improve EBV accuracy for survival time in layers showing cannibalism.

We considered four DGE-IGE models to predict survival time: one model analysing survival time and three models analysing survival as a repeated binomial trait. It was also tested whether EBV were improved by including timing of IGE expression in the analyses. Approximate EBV accuracies were calculated by cross-validation. The data contained 13 repeated survival records of two White Leghorn layer lines W1 and WB.

Compared to analysing survival time, EBV accuracy was reduced when timing of IGE expression was included in the model. EBV accuracy was higher when repeated measures were used. Using repeated measures instead increased EBV accuracy by 10 to 21% for W1 and 2 to 12% for WB.

Our results suggest that EBV accuracy for survival time in laying hens can be improved using repeated measures models. This is important because more accurate EBV contribute to increased selection response.

3:05 – 3:35 PM Coffee break (30 min)

3:35 – 3:50 PM Selected Poster Presentation: Predicting causal variants affecting expression using whole genome sequence and RNA-seq from multiple human tissues

Dr. Andrew A Brown[1], Dr Olivier Delaneau[1], Professor Emmanouil Dermitzakis
[1]*University of Geneva*

Genome-wide association studies have produced thousands of genetic regions associated with traits, but identifying the particular variant responsible has proved more difficult. Systematically identifying causal variants needs full ascertainment of all SNPs within a region, which is possible with whole genome sequencing but current studies are limited by cost. However, this is feasible when studying gene expression, as relatively small sample sizes are sufficient to provide thousands of associations. We present work which has combined RNA-seq data from 4 tissues and whole genome sequence for 520 individuals. We show that whole genome sequence does not dramatically increase power to find associations relative to genotype data from arrays (26,351 eQTL discovered compared to 27,659 with sequence). However, sequence eQTL were more enriched in all relevant tissue open chromatin regions called by the Roadmap Epigenomics Project (with odds ratios 1.2 to 1.4, P = 2.6e-3 to 1e-10, 30/31 Bonferroni significant). This suggests that with sequence data, the causal variant is more often identified. Building on these results, we developed a method (CaVEMaN), based on non-parametric resampling, to assess the likelihood of a specific variant being causal. We compare this to other methods, such as considering the difference between peak and second association. In simulated data CaVEMaN provides a slightly conservative estimate of the probability that a variant is causal. In addition, CaVEMaN more often predicts whether the peak eQTL is causal than other methods, as it accounts for phenotypic variability and full linkage structure in the region. This second result was confirmed independently using real data; the top 10% of eQTL predicted to be causal by CaVEMaN were enriched in open chromatin regions relative to those ranked by difference in top P values (OR 1.01 – 2.00, 7/31 Bonferroni significant).

3:50 – 4:30 PM Using human genetic variation for causal inference: methods, opportunities, and challenges for Mendelian randomization

Dr. Brandon Pierce[1]
[1]*Departments of Public Health Sciences and Human Genetics, University of Chicago*

"Mendelian randomization" (MR) is a method for estimating the causal effect of a disease risk factor on a health outcome using inherited genetic variation as an instrumental variable. Use of the MR

method for causal inference has grown enormously in popularity over the past five years. This growth was enabled by the rapid increase in knowledge regarding genetic variants that influences disease risk factors of interest (e.g., biomarkers, physical traits, behaviors) due to the past decade of genome-wide association studies (GWAS). This talk will first provide an overview of the fundamentals of MR estimation, including single-sample and two-sample methods, the use of summarized association estimates, as well as multi-SNP and whole-genome instrumental variables. I will then describe emerging opportunities for MR research presented by the wealth of knowledge gained from GWAS and the emergence of large-scale GWAS consortium and genotyped population-based cohorts. Recent MR work from the GAME-ON (Genetic Associations and Mechanisms in Oncology) post-GWAS consortium will be described. Finally, I will describe challenges associated with obtaining unbiased MR estimates, most notably, the assessment of potential violations of the assumptions required for MR estimates to be valid estimates of causal relationships.

4:30 – 6:30 PM Poster Session and Social Hour

WEDNESDAY, JUNE 15

8:30 AM – 12:05 PM SESSION V:

POPULATION AND EVOLUTIONARY GENETICS IN THE GENOMICS ERA

8:30 – 8:40 AM Chair Remarks: Dr. Patrick Phillips, University of Oregon, USA

8:40 – 9:20 AM Heterochromatin diversity and quantitative trait evolution

Dr. Andrew Clark[1]
[1]*Cornell University*

Substantial portions of the genomes of higher organisms consist of tandem repeats of simple sequences. The bulk of these are located near centromeres in regions called centromeric heterochromatin. Analysis of kmer frequencies in raw read data allow us to build dictionaries of relative abundances of these repeats and to quantify variation among species and individuals. Lines of Drosophila melanogaster vary widely in heterochromatin abundance, and an easy way to perturb heterochromatin abundance in a genome is to use Y chromosome replacement lines. We characterize the associations between kmer repeats and genome-wide gene expression changes, and begin to explore how Y-linked heterochromatin perturbs chromatin state genome-wide. Chromatin modulation is yet another source of variation that must play a role in quantitative trait evolution.

9:20 – 10:00 AM Genetics of the island rule

Dr. Bret Payseur[1]
[1]*University of Wisconsin-Madison*

Repeated patterns of diversification offer insights into general mechanisms of evolution. A striking example is the tendency of organisms on islands to evolve exceptional body sizes. To understand the genetic basis of this intriguing phenomenon, we study the largest wild house mice in the world, which recently colonized the remote Gough Island. Using comparisons to a mainland strain, we demonstrate that Gough Island mice achieve their remarkable body size by accelerating growth early in life. We integrate genetic mapping of weight and skeletal phenotypes, genome-wide measurement of gene expression in key metabolic organs, and genome sequencing to provide a rare genetic portrait of extreme body size evolution on islands. We use population genomics to reconstruct the history of these unique mice, including natural selection targeting candidate genes for size evolution.

10:00 – 10:30 AM Coffee break (30 min)

10:30 – 10:45 AM Selected Poster Presentation: Genome-wide and species-wide variation in C. elegans reveals association of telomere length with population differences in pot-2

Daniel E. Cook[1,2], Stefan Zdraljevic[1,2], Robyn E. Tanny[1], Erik C. Andersen[1,2]
[1]*Dept. of Molecular Biosciences, Northwestern University,* [2]*Interdisciplinary Biological Sciences Program, Northwestern University*

Model organisms have greatly advanced our understanding of biological phenomena. However, experiments in model organisms are frequently carried out in a single laboratory strain (e.g. N2 in C. elegans). This approach, although useful for many applications, neglects insights that can be made by examination of genetic diversity in natural populations. This largely untapped resource can help tease apart the genetic factors responsible for complex traits and address questions concerning etiology of disease, genome evolution, and population genetics.

Given the benefits of studying natural populations, we have sequenced the whole genomes of 211 wild isolate strains of C. elegans that represent 152 distinct genome-wide haplotypes. To date, this collection is the largest set of wild isolate sequences in C. elegans. Sequences were aligned and achieved a median coverage of 72x, giving us the ability to identify rare and common variants with high confidence. Across all of these strains, we called approximately 1.3 million single nucleotide variants that are predictive of functional changes in C. elegans.

Using this dataset, we mapped trait differences ascertainable from sequence data, including estimates of telomere length. We found that telomere length is strongly associated with a region on chromosome II that contains the gene pot-2 (Protection of Telomeres). POT-2 represses the telomere-lengthening enzyme telomerase. We identified a nonsynonymous (F68I) variant within the OB-fold domain of POT-2 that is correlated with longer telomeres. Additional analysis on mutagenized strains revealed enrichment for mutations specifically within the OB-fold of POT-2 in long-telomere strains. We also examined whether telomere length affected brood size, longevity, or the germlines of strains and did not observe any correlations. Our results suggest that telomere length does not correlate with fitness. Additionally, the F68I variant is not under selection in wild populations, indicating that variation in telomere length among individuals likely has little appreciable benefit.

10:45 – 11:25 AM Good dad, bad dad: the genetic basis of parental care

Dr. Andres Bendesky[1], Young-Mi Kwon[1], Dr. Jean-Marc Lassance[1], Caitlin Lewarch[1], Dr. Shenqin Yao[1], Dr. Brant Peterson[1], Meng-Xiao He[1], Dr. Catherine Dulac[1], Dr. Hopi Hoekstra[1]
[1]*HHMI/Harvard*

Parental care is essential for the survival of mammals, yet the mechanisms underlying its evolution remain largely unknown. Here we show that two closely related species of Peromyscus mice, P. polionotus and P. maniculatus, differ greatly in parental behavior and that these differences are heritable. Using a quantitative genetic approach, we identify 11 genomic regions that affect parental care, eight of which have sex-specific effects, suggesting that parental care can evolve through independent genetic mechanisms in males and females. Furthermore, some regions affect parental care broadly whereas others affect specific aspects of parental behavior, such as nest building. Transcriptome analysis shows that, of the genes that reside in genomic regions linked to nest-building behavior, vasopressin is strongly differentially expressed in the hypothalamus of the two species, with increased levels associated with less nest building. Using chemogenetics in Mus musculus, we show that we can promote nest building by inhibiting hypothalamic vasopressin neurons and decrease nest building by exciting these neurons. Together, our results define the genetic properties of parental behavior evolution in Peromyscus and indicate that variation in an evolutionary ancient neuropeptide contributes to interspecific differences in parental care.

11:25 – 12:05 PM Genetics of adaptation to changing climate in Arabidopsis thaliana

Dr. Johanna Schmitt[1], Dr. Alexandre Fournier-Level[2]
[1]*University of California, Davis,* [2]*University of Melbourne*

Whether and how natural populations will adapt fast enough to persist in the face of rapid climate change is a critical question for 21st century evolutionary biology. The annual plant Arabidopsis thaliana provides a model system for investigating and predicting patterns of climate adaptation. Extensive genomic data exist from many accessions collected across the species' climate range. These data can be combined with phenotypic data from common garden experiments using genome-wide association analysis to identify traits and loci associated with real-time fitness in different climates, and to test for a geographic signature of adaptation to climate in the site of origin. Genetic resources such as mutants and mapping populations make it possible to examine how perturbation of specific developmental pathways will affect life history expression in dynamic real-world environments. Experimental data can be used to inform and parameterize models to predict life history responses to climate change. In particular, genomic selection methods combined with simulation of evolutionary trajectories make it possible to explore the dynamics of phenotypic and genotypic evolution in replicated scenarios of climate change. Using this approach, we simulated evolutionary trajectories of populations under different scenarios of warming climate and poleward migration. The genetic basis of adaptation and dynamics of evolutionary change differed qualitatively among scenarios. Populations with genetic diversity reduced by prior selection adapted less well to novel conditions, demonstrating that adaptation to rapid climate change requires the maintenance of sufficient standing variation.

12:05 - 5:00 PM Afternoon Social Activities (see "Things to do, places to see" on page 65.)

5:00 - 7:00 PM Happy Hour – Monona Terrace Rooftop Center Circle

THURSDAY, JUNE 16

8:30 AM – 12:05 PM SESSION VI:

BIG DATA IN GENOMICS

8:30 – 8:40 AM Chair Remarks: Dr. Karl Broman, UW-Madison, USA

8:40 – 9:20 AM False Discovery Rates - a new deal

Dr. Matthew Stephens[1]
[1]*University of Chicago*

False Discovery Rate (FDR) methodology, first put forward by Benjamini and Hochberg, and further developed by many authors - including Storey, Tibshirani, and Efron - is now one of the most widely used statistical methods in genomics, among other areas of application. A typical genomics workflow consists of i) estimating thousands of effects, and their associated p values; ii) feeding these p values to software (e.g. the widely used qvalue package) to estimate the FDR for any given significance threshold. In this talk we take a fresh look at this problem, and highlight two deficiencies of this standard pipeline that we believe could be improved. First, current methods, being based directly on p values (or z scores), fail to fully account for the fact that some measurements are more precise than others. Second, current methods assume that the least significant p values (those near 1) are all null - something that initially appears intuitive, but will not necessarily hold in practice. We suggest simple approaches to address both issues, and demonstrate the potential for these methods to increase the

number of discoveries at a given FDR threshold. We also discuss the connection between this problem and shrinkage estimation, and problems involving sparsity more generally.

9:20 – 10:00 AM Predicting the impact of rare regulatory variation

Dr. Alexis Battle[1]
[1]*Department of Computer Science, Johns Hopkins University*

The enormous increase in availability of full human genomic sequences presents great opportunity for understanding the impact of rare genetic variants. Based on current knowledge, however, we are still limited in our ability to interpret or predict regulatory consequences of rare variants. Association methods are inherently limited for variants observed in only one or a few individuals. Diverse genomic annotations such as growing resources of epigenetic data and have been shown to be informative regarding regulatory elements, but are only moderately predictive of impact for individual variants. The availability of personal RNA-seq and other cellular measurements for the same individuals with genome sequencing offers a new avenue for integrated methods for prioritizing rare functional variants. By considering informative genomic annotations along with molecular phenotyping, we are able to identify the regulatory impact of rare genetic variants largely excluded from previous analyses. We have evaluated the impact of rare regulatory variants using whole genome sequences and corresponding RNA-sequence in 44 different tissues samples from the GTEx project and report. We demonstrate that rare variants in promoter elements, conserved regions, and variants that affect splicing are associated with extreme changes in gene expression across multiple tissues. Additionally, we have developed a Bayesian machine learning approach that integrates whole genome sequencing with RNA-seq data from the same individual, leveraging gene expression levels along with diverse genomic annotations and performing joint inference to identify likely functional rare regulatory variants. We have applied this model to the GTEx data to prioritize the most likely functional rare regulatory variants for each individual. We demonstrate that integrative models perform better than predictions from DNA-sequencing or RNA-sequencing alone. Our probabilistic model of rare regulatory genetic variants offers potential for identifying deleterious regulatory variants from individual genomes.

10:00 – 10:30 AM Coffee break (30 min)

10:30 – 10:45 AM Selected Poster Presentation: Use of electronic health record to predict family relationships for genetic epidemiology research

Mr. Xiayuan Huang[1], Dr. Robert Elston[3], Dr. Guilherme Rosa[1], Mr. John Mayer[2], Dr. Ye Zhan[2], Dr. Terrie Kitchner[2], Dr. Page David[1], Dr. Scott Hebbring[1,2]
[1]*University of Wisconsin-Madison,* [2]*Marshfield Clinic Research Foundation,* [3]*Case Western Reserve University*

Diseases are the result of genetic and/or environmental factors. Understanding the etiology of human disease is necessary to advance precision medicine. Pedigree analysis is a longstanding approach to gain insight into the underlying genetic factors in human health. In the last decade, pedigree analysis has largely given way to large population-based studies on samples of independent persons as a result of high-throughput genotyping methodologies and the ability to collect large numbers of unrelated cases and controls. A reverse paradigm shift could occur if the construction and use of family pedigrees could be combined with classical genetic analyses in a high-throughput manner. The goal of this study is to demonstrate that family pedigrees may be rapidly predicted from an electronic health record (EHR) system with no human intervention. Results demonstrate our ability to identify with high probability 173,368 family pedigrees using basic demographic data available in most EHRs. We further show as proof-of-principle that these large patient pedigrees may be applied to genetic research. With the anticipated widespread application of genomic medicine, in combination with methods capable of predicting family pedigrees linked to extensive and longitudinal

phenotypic data, a revolution may occur in how human genetic research is conducted for the advancement of precision medicine.

10:45 – 11:25 AM Sequence based virtual breeding: A tool to evaluate GWAS and genomic selection in massive datasets

Miguel Perez-Enciso[1,2], N. Forneris[1], G. de los Campos[3], A. Legarra[4]
[1]Centre for Research in Agrigenomics, [2]Institut Català de Recerca i Estudis Avançats (ICREA), [3]Michigan State University, [4]INRA

Extant simulation tools such as coalescence and gene dropping that incorporate mutation are unable to simulate complete genomes. In an era when massive investments in sequencing are being made, powerful realistic simulators are fundamental to optimize experimental designs and interpret their results. In this study we introduce software that simulated complete genomes from real sequence-derived SNP data and performs gene dropping in an extremely efficient way. It implements any number of traits (binary or continuous) or QTNs, dominance, two-locus epistasis, variable recombination rates along the genome, sex chromosomes, and selection.

In the first experiment, we explored how different SNP sets performed for genomic selection purposes using 107 public pig sequences and a simulated pedigree of 3,400 individuals (~28M SNPs). The quantitative trait was determined by ca. 1,000 SNPs that were chosen within those genes with largest Fst between wild boar and domestics, lowest Tajima's D and predicted to have at least a moderate effect in the protein sequence. In a wide variety of scenarios that comprised outbred and crossbred populations, we found negligible differences in predictive accuracy between using complete sequence, Illumina's 60k, Affymetrix 700k or 500k random SNPs, when a Bayesian GBLUP procedure was implemented.

For the second experiment, we downloaded the 205 sequences from the Drosophila Genome Reference Panel 2 (http://dgrp2.gnets.ncsu.edu/, ~4M SNPs). We performed 8 generations of divergent selection, with and without complementary epistasis. In both architectures, the most associated SNPs and their effects as identified by Ober et al. (10.1371/journal.pone.0126880) for chill recovery were employed. If present, epistasis accounted for ~10% of phenotypic variance. Genomic selection was performed using all SNPs available up to the current generation, based on a purely additive model in all cases. We found no difference in response between the additive and the epistatic plus additive architectures. A strong asymmetric response was observed, though.

11:25 – 12:05 PM A new highly scalable method for calculating probabilities of fitness consequences for point mutations across the human genome

Prof. Adam Siepel[1]
[1]Simons Center for Quantitative Biology, CSHL

A major focus of my research program is to characterize the influence of natural selection on noncoding elements in mammalian genomes, based on patterns of genetic variation within and between species. In this talk, I will first review a probabilistic framework we have developed, called INSIGHT, that measures the influence of selection on collections of short, interspersed noncoding elements across the genome. Next, I will describe how we have combined the INSIGHT framework with large-scale functional genomic data, to estimate the probability that a point mutation at each position in a genome will influence fitness. These fitness consequence (fitCons) scores serve as evolution-based measures of potential genomic function, and, compared with conventional conservation scores, they show considerably improved prediction power for known cis-regulatory elements. Finally, I will describe a recent extension of these methods, called L-INSIGHT (linear-INSIGHT), that allows fitCons scores to be computed based on much larger collections of genomic covariates. L-INSIGHT combines probabilistic evolutionary modeling with a linear-logistic model for genomic covariates, allowing for greatly improved scaling properties and the ability to jointly analyze

massive functional and comparative genomic data sets. Preliminary results indicate that these new L-INSIGHT scores have substantially improved genomic resolution and predictive power. We are in the process of building fitCons maps for several species, including human, Arabidopsis and rice.

12:05 – 1:20 PM Lunch

1:20 – 4:30 PM SESSION VII:
EPISTASIS AND G X E INTERACTION

1:20 – 1:30 PM Chair Remarks: Dr. Lucia Gutierrez, UW-Madison, USA

1:30 – 2:10 PM Modeling genotypic responses across environments: the "problem" of genotype by environment interactions

Dr. Marcos Malosetti[1]
[1]*Wageningen University*

Phenotypes result from genetic information and external environmental stimuli integrated over developmental time. Genotype by environment interactions occur because different genotypes do not necessarily react in the exact same way to equal conditions. Differential plasticity, adaptation, stability, and resilience are therefore part of the problem researchers have to deal with in multi-environment data. While Gene-Environment systems are clearly high-dimensional and non-linear, linear and bilinear statistical models have proved successful to study phenotypic responses over multiple environments. Beyond answering fundamental questions, an important driver of the modeling efforts is the prediction of genotype performance over a given environmental space. A common feature of all these models, is that they depart from the classical additive model y = G + E + GxE towards models consisting of genotype-specific and/or environment-specific parameters (separability). In particular, models that incorporate explicit genotypic and/or environmental information have great potential (factorial regression models, QTLxE models, GE-prediction models). In this presentation an overview of different models will be given and discussed with applications in plants. Also the integration with other strategies to model genotypic responses, for example, by incorporating crop growth models, will be discussed. The underpinning idea of the presentation is that the integration of different sources of information and modeling approaches can contribute to alleviate the "problem" of genotype by environment interactions and even, turn it into an opportunity.

2:10 – 2:50 PM How to define and estimate genomic epistasis consistently?

Dr. Zhao-Bang Zeng[1]
[1]*Bioinformatics Research Center, Departments of Statistics and Biological Scineces, North Carolina State University*

In quantitative genetics we partition genetic variance into additive, dominant and epistatic variances. How to define and estimate those variances, particularly epistatic variance, unbiasedly and reliably is a challenge. GBLUP (genomic best linear unbiased prediction) has been used to estimate genome-wide epistatic variance using dense markers. By using an orthogonal genetic model, we show that estimates of epistatic variance by GBLUP can be made consistent in full and reduced models, and unbiased. This is demonstrated through simulation evaluation. We applied the method to the maize NAM population (McMullen et al. 2009, Science) and the rice "immortalized F2" population (Hua et al. 2003 PNAS; Zhou et al. 2012 PNAS) to assess the magnitude of epistatic variance. For the NAM population, the additive variance generally accounts for 10 to 50% of the total variance for a variety of traits, and the epistatic variance accounts for additional 10 to 15% of the total variance. The results confirm that many life history traits that have relatively lower additive variance tend to have

relatively large amount of epistatic variance. For the "immortalized F2" population, grain yield has very significant amount of additive x additive, additive x dominant and dominant x dominant epistatic variances. On epistasis, there is actually a more serious issue about how to define a genetic model that is consistent with respect to partition of genetic effects and genetic variances in different number of QTL and for the whole genome. Unfortunately, the traditional quantitative genetic model is not consistent in this respect. This led to long-standing separation of genetic effects from genetic variances, and inconsistence in estimation of different genetic variance components in the full and reduced models by using the traditional genetic model. In the era of genomics, it is the time now to realize the importance in promoting the use of a genetic model that is multiple locus consistent.

2:50 – 3:05 PM Selected Poster Presentation: The epistasis boundary: Linear vs. nonlinear genotype-phenotype relationships

Dr. Serge Sverdlov[1], Prof. Elizabeth A. Thompson[1]
[1]*University of Washington*

Nonlinearity in the genotype-phenotype relationship can take the form of dominance, epistasis, or gene-environment interaction. The importance of nonlinearity, especially epistasis, in real complex traits is controversial. Network models in systems biology are typically highly nonlinear, yet the predictive power of linear quantitative genetic models is rarely improved by addition of nonlinear terms, and association studies detect few strong gene-gene interactions. We find that complex traits satisfying certain conditions can be well represented by a linear genetic model on an appropriate scale despite underlying biological complexity. Recent mathematical results in separability theory determine these conditions, which correspond to three biological criteria (Directional Consistency, Environmental Compensability, and Pathway Redundancy) together making up an Epistatic Boundary between systems suitable and unsuitable for linear modeling.

For genetic traits controlled by limited numbers of loci and alleles, our algorithm enumerates all possible trait structures and finds exact or error-minimizing linearizing transformations by formulating a constrained optimization program. We find that the fraction of possible distinct genetic traits satisfying simple criteria that can be fully or approximately linearized is high for small systems and falls with system complexity.

For nonlinear traits, we introduce a combinatorial classification of types of nonlinearity in a network context. We use this to illustrate how upstream controlling genes, potentially more important to explaining biological function, can be intrinsically harder to detect by GWAS than their downstream controlled counterparts.

3:05 – 3:35 PM Coffee break (30 min)

3:35 – 3:50 PM Selected Poster Presentation: Investigation of genotype by environment barriers to maize adaptation using near-isogenic lines capturing an allelic series

David M Wills[1], Nicholas Lauter[2], Teclemariam Weldekidan[3], Miriam Lopez[2], Natalia de Leon[4], Sherry A Flint-Garcia[1], James B Holland[5], Seth C Murray[6], Wenwei Xu[7], Randall J Wisser[3]
[1]*USDA-ARS Plant Genetics Research Unit, University of Missouri,* [2]*USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University,* [3]*Department of Plant and Soil Sciences, University of Delaware,* [4]*Department of Agronomy, University of Wisconsin,* [5]*USDA-ARS Plant Science Research Unit, North Carolina State University,* [6]*Department of Soil and Crop Sciences, Texas A&M University,* [7]*Texas A&M Agrilife Research*

Flowering time can act as a strong barrier to the adaptation of a crop to a new environment. For tropical maize, flowering time is typically delayed in long day environments. This photoperiod response may mask the effects of beneficial alleles via pleiotropy, and the selection required for flowering time adaptation may result in a loss of beneficial alleles in coupling phase with photoperiod sensitive haplotypes. To address questions about the nature of pleiotropy, genotype by environment

(GxE) interaction, linkage drag and allelic variation in terms of environmental adaptation, we created a population design constructed as a set of Near-Isogenic Lines capturing an Allelic Series (NILAS). A NILAS is a collection of inbred lines differing at a single specific genomic locus for a range of functional alleles. Marker-assisted selection was used to construct NILAS for each of four photoperiod quantitative trait loci (QTL) implicated as primary barriers to maize adaptation to temperate environments. Fourteen distinct tropical x temperate NILAS contrasts were created using seven tropical inbred lines as allele donors and two different temperate inbred lines as genetic backgrounds. Each NILAS captured 12 overlapping introgressions tiled across the QTL region. The NILAS were assessed in eight environments extending from approximately 18° N latitude (Puerto Rico) to approximately 43° N latitude (Wisconsin) in two replicates of a four-way split plot design. We present evidence for an allele series, GxE, phenological pleiotropy, and linked variation based on the evaluation of >20 characteristics. The NILAS is a powerful design to phenotypically, genetically, and ecologically characterize genomic loci. Our findings demonstrate the complexity of quantitative variation within a chromosome based on multi-environment trials and how the NILAS design is extensible for continued studies.

3:50 – 4:30 PM Something about epistasis

Dr William Hill[1]
[1]*University of Edinburgh*

There is some controversy in the importance of epistasis to genetic analysis of quantitative traits. Genes act within and across pathways and some GWAS studies have identified many epistatic QTL. Little of the variance is, however, either expected or generally found to be epistatic in large outbred populations or, as Griffing showed long ago, is expected to contribute to cumulative selection response and be worth considering in genomic prediction models. Recent analysis by Barton and colleagues (in PNAS) using an infinitesimal model, however, suggest that in finite populations epistatic variance can contribute to long term response. I shall review and try to reconcile these various strands.

4:30 – 6:30 PM Poster Session and Social Hour

FRIDAY, JUNE 17

8:30 AM – 12:05 PM SESSION VIII:

GENOMIC INFORMATION IN PREDICTION

8:30 – 8:40 AM Chair Remarks: Dr. Ann Stapleton, University of North Carolina Wilmington, USA

8:40 – 9:20 AM Improving genomic predictions with sequence data and biological information

Dr. Ben Hayes[1,2], Hans Daetwyler[2,3], Amanda Chamberlain[2], Iona Macleod[2,4], Kathryn Kemper[4], Mike Goddard and the 1000 bull genomes consortium[4]
[1]*Queensland Alliance for Agriculture and Food Innovation, University of Queensland,* [2]*Department of Economic Development, Jobs, Transport & Resources, AgriBio,* [3]*La Trobe University, AgriBio,* [4]*4Faculty of Veterinary and Agricultural Sciences, University of Melbourne*

The identification of causal mutations which affect complex traits in cattle could improve the accuracy of genomic estimated breeding values, particularly across breeds, and with greater persistency of accuracy across time. 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, including 64.8

million SNP and 2.5 million indel have been identified. The challenge is now to determine which of these millions of variants are the causal mutations underlying variation in the complex traits, and then to use these variants in genomic predictions. This challenge is magnified by the fact that the size of effects of such causal mutations are likely to be small, given the large number of mutations typically affecting complex traits. We address this challenge with three strategies. The first is development of new highly computationally efficient genomic prediction algorithms, able to handle millions of variants genotyped in tens of thousands of animals. The second strategy is to use intermediate phenotypes, where the effect of the mutation is much larger than on the complex trait phenotype, such as gene expression and protein abundance. The third strategy is to incorporate biological information, such as genome annotations, into the genomic predictions. Several examples of using these strategies to identify potential causal mutations affecting protein content of milk from dairy cattle are given. The results do highlight the need for better annotation of the bovine genome – many of the most significant mutations are in poorly annotated genomic regions, likely regions that regulate gene expression. This situation will be improved in the near future, with efforts such as the functional annotation of animal genomes (FAANG) consortium.

9:20 – 10:00 AM Next generation animal and plant breeding

Dr. John Hickey[1], Mara Battagin[1], Janez Jenko[1], Roberto Antolin[1], Daniel Money[1], Adriana Somavilla[1], Roger Ros Freixedes[1], Mick Watson[1], C. Bruce A. Whitelaw[1], John A. Woolliams[1], Matthew A. Cleveland[2], Alan J. Mileham[2], William Herring[2], Andreas Kranis[3], R. Chris Gaynor[1], Mackay Ian[4], Alison Bentley[4], Gregor Gorjanc[1]
[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, [2]Genus/PIC PLC, [3]Aviagen Limited, [4]John Bingham Laboratory, NIAB

Genomic information has revolutionized animal breeding programs over the past decade and is increasingly common in plant breeding programs. Genomic information is also causing the tools and techniques that can be used in plant and animal breeding to coalesce. In this contribution the current status of genomic selection in plant and animal breeding programs is reviewed and next steps in each sector are proposed. In livestock breeding programs the proposed next step is termed Genomic Selection 2.0 in which sequence information can be generated on millions of individuals cost effectively to enable new breeding tools, including the promotion of alleles by genome editing (PAGE) and manipulation of recombination rate, be used to drive increased genetic gain. In plant breeding programs the proposed next step is termed Genomic Selection 1.0 which would be utilized within completely reorganized breeding program designs. This reorganization would involve explicit splitting of the breeding program into population improvement and product development stages. Within the population improvement classical animal breeding tools such as economic selection indices and optimal contributions could have greater impact and much greater and more holistic rates of genetic progress could be achieved by the entire program.

10:00 – 10:30 AM Coffee break (30 min)

10:30 – 10:45 AM Selected Poster Presentation: Genomic prediction based on interaction networks of markers: How to incorporate prior experimental information

Dr. Johannes WR Martini[1], Dr. Valentin Wimmer[2], Prof. Dr. Henner Simianer[1]
[1]Georg-August University, [2]KWS SAAT SE

Different approaches have been developed in quantitative genetics to model non-additive relationships of individuals. Among them, the epistasis marker effect model which assumes that the genetic value is not only the sum of additive effects, but additionally of pairwise interactions of genetic markers, has been discussed in the last years. Dealing with the full epistasis model with all pairwise interactions, two characteristics attract attention. One obvious fact is that the number of

variables in the model increases drastically from the number of markers p in the purely additive model to approximately p^2/2 in the full epistasis model. Since the model also incorporates a large number of potentially unimportant variables, the question arises whether predictive ability can benefit from variable selection. Another important point is the coding-dependent performance of the epistasis model in genomic prediction, raising the question of how to code the markers optimally for a given data set. We show that both characteristics offer possibilities for incorporating knowledge from previous experiments into the prediction method: Instead of modeling all pairwise interactions, prior knowledge from previous experiments can be used to infer interaction subnetworks by restricting the model to biologically relevant interactions. These subnetworks which may very roughly reflect genetic architecture of the trait can afterwards be used for other data sets as prior information. Moreover, the coding-dependent performance of the epistasis model allows for adapting the coding to previous results which can afterwards be used as prior information in new experiments. The presented results illustrate that properties of the epistasis model which may at first seem unattractive can also provide the flexibility to reflect genetic architecture in at least two ways.

10:45 – 11:25 AM Model training across multiple breeding cycles significantly improves genomic prediction accuracy in hybrid rye breeding

Chris-Carolin Schön[1], Christina Lehermeier[1], Eva Bauer[1], Hans-Jürgen Auinger[1], Manfred Schönleben[1], Malthe Schmidt[2], Viktor Korzun[2], Anres Gordillo[2], Peer Wilde[2]
[1]Technical University of Munich, Plant Breeding, [2]KWS LOCHOW GmbH

In many hybrid breeding programs, application of genome-based prediction is expected to increase selection gain because of long selection cycles in population improvement and development of hybrid components. Essentially two prediction scenarios arise: i) prediction of the genetic value of individuals from the same breeding cycle in which model training is performed and ii) prediction of individuals from subsequent cycles. It is the latter from which a reduction in cycle length and consequently the strongest impact on selection gain is expected. Based on experimental data from multiple selection cycles of a hybrid rye breeding program we investigate the efficiency of genome-based prediction for agronomic traits within and across selection cycles. The performance gap between genomic and pedigree-based best linear unbiased prediction (GBLUP and PBLUP) increases significantly when model calibration is performed on aggregated data from several cycles. If selection cycles are connected by a sufficient number of common ancestors and prediction accuracy has not reached a plateau with respect to sample size in model training, aggregating data from several preceding cycles is recommended despite decreasing relatedness over time. We will demonstrate the effect of the mating design on across-cycle prediction accuracies and discuss the impact of integrating results from genome-wide association studies into the classical GBLUP framework. Insights derived from this study will be used as a starting point for evaluating optimum allocation of resources when integrating genome-based prediction in multi-stage selection schemes.

11:25 – 12:05 PM Genomic prediction in forest tree breeding: advances and challenges in tropical Eucalyptus

Dr. Dario Grattapaglia[1,2]
[1]EMBRAPA Genetic Resources and Biotechnology, [2]Genomic Science Program, Universidade Católica de Brasília

The challenge of tree improvement programs is the long time interval between the breeding investment and the deployment of improved genetic stock. Notwithstanding important advances in QTL mapping and association genetics in forest trees, genomics has not had any significant impact in operational tree breeding. Reasons include the limitations of early genomic technologies, the genetic heterogeneity of undomesticated trees but mainly the overoptimistic view of the architecture of complex traits. The advent of high-throughput genotyping technologies, coupled to Genomic Selection (GS), provide a new paradigm to integrate genomics into breeding. GS captures most of the

genome-wide effects that linkage and association genetics fail to detect. The milestone publication of the Eucalyptus genome allowed us to develop a multi-species 60KSNP chip based on whole-genome resequencing of 240 Eucalyptus tree genomes, providing high-density genotyping across all ten planted Eucalyptus species worldwide. Using this genotyping platform in a number of breeding programs in Brazil, we have shown that GS prediction can match phenotypic selection for growth and wood quality traits. GS may significantly reduce the length of a breeding cycle by ultra-early selection of multi-trait ranked seedlings, precluding progeny trials. Since predictive abilities are impacted by GxE and driven mainly by relatedness between 'training' and 'validation', we expect that population-specific predictive models will be needed. Besides the lack of experimental data on validation across generations, open questions still include the application of GS to hybrid breeding and the use of model updating to counterbalance decay of relationship and LD while adjusting to the rapid environmental changes expected to have greater consequence on the continued performance of GS in trees. Logistics and cost are also challenges faced to fully integrate this new breeding technology into routine tree improvement. We expect however that GS will finally make genomics have a true impact in tree breeding.

12:05 – 1:20 PM Lunch

1:20 – 4:00 PM SESSION IX:
TRANSLATIONAL QUANTITATIVE GENOMICS

1:20 – 1:30 PM Chair Remarks: Dr. Lucia Albuquerque, Sao Paulo State University, Brazil

1:30 – 2:10 PM Prospects for accurate risk prediction of common diseases and disorders

Dr. Naomi Wray[1]
[1]*Queensland Brain Institute, The University of Queensland*

The motivation for prediction of disease is its potential impact on prevention, diagnosis and treatment, but the extent to which these goals hold true will vary between diseases. For example, accurate genetic risk prediction for Huntington's Disease, a Mendelian disorder, has been available for over 20 years, yet this opportunity is often not followed through by family members since there is no actionable outcome. In contrast, for common complex genetic diseases, while a genetic predictor alone can never be accurate for an individual, genetic risk stratification could have clinical utility. Risk stratification could identify a high risk class that includes the majority of those who will become affected in their lifetime (high sensitivity) even though the majority of those in this high risk class will not be affected (poor specificity). However, utility is less clear when diagnostic interest is in stratifying amongst those presenting at a clinic with non-specific symptoms. Although the methodology of genetic risk prediction of human disease parallels genetic evaluation in agriculture fundamental differences reflect the data structure available for generating and validating predictors, and that the prediction goal is of an uncommon binary phenotype of an individual, compared to the mean value of a quantitative trait in the next generation. Accurate prediction of a phenotype, requires the genetic predictor to be enhanced to include non-genetic risk factors. The genomics era allows the inclusion genomic biomarkers, which could reflect the downstream consequences disease and of non-measured environmental risk factors. While disease risk prediction is likely to some have clinical utility, in the era of personal or precision medicine it is important to not to oversell what can be realistically achieved.

2:10 – 2:50 PM The application of genomics to genetic improvement of livestock

Dr. Jack CM Dekkers[1]
[1]*Iowa State University*

While the use of individual genetic markers through marker-assisted selection saw only limited adoption, the livestock breeding industry has now fully embraced the use of molecular genetic information in the form of genomic selection, in particular in the dairy cattle breeding industry, where it has led to a paradigm shift in selection programs and selection strategies. For various reasons, which will be discussed, the use of genomic selection and prediction in other livestock species has been less 'revolutionary'. The purpose of this presentation is to give an overview of the current state of the use of genomic prediction and genomic selection in animal breeding programs, as well as an overview of the successes, limitations, and challenges, and future opportunities for the use of genomics in genetic improvement of livestock. Opportunities and challenges for the use of genomic prediction to select for 'difficult' traits, will be discussed, including crossbred performance, disease resistance, and other hard-to-measure traits. Breeding strategies to better capitalize on advances in genomics while controlling rates of inbreeding will be discussed also.

2:50 – 3:20 PM Coffee break (30 min)

3:20 – 4:00 PM Translational quantitative genomics in plant improvement

Dr. Rex Bernardo[1]
[1]*University of Minnesota*

Forty years ago, during the first International Conference on Quantitative Genetics held at Iowa State University in August 1976, Professor R.C. Lewontin described the conference subject as follows: "Quantitative genetics, upon which modern plant and animal breeding is based, is an attempt to produce knowledge by a systematization of ignorance." This description was rooted in the dearth of knowledge at that time about quantitative trait loci (QTL): their number, location, allele frequencies, phenotypic effects, interlocus interactions, and responses to environments. Developments in the last quarter century have led to greatly increased knowledge about the genome. However, the biochemical, physiological, and developmental paths from genes to grain yield, fruit quality, or resistance to biotic and abiotic stresses remain a black box. Routine utilization of genomic information in plant improvement has been mostly limited to the use of single nucleotide polymorphism (SNP) markers in crossing good by good and selecting the best. Marker-trait linkages are currently utilized for traits, such as disease resistance, that might have major QTL; genomewide predictions are currently utilized for traits, such as yield, for which major QTL are absent. These approaches focus on the G component of the P = G + E equation. Advanced phenomics (P), envirotyping (E), and genome editing (another G) are more recent developments. Translational quantitative genomics in plants has focused mainly on selection, but much more emphasis is needed on the use of genomics tools to find and characterize new and useful genes and germplasm for complex traits.

4:00 – 4:40 PM CLOSING SESSION

4:00 – 4:40 PM Recognition ceremony (chaired by Dr. Bill Hill) and Closings (Drs. Natalia de Leon and Guilherme Rosa, ICQG5 co-chairs)

6:30 - 11:00 PM Conference Dinner (optional – tickets required). This event will be held at the Overture Center.

MONDAY POSTERS

Poster Number	Title	Presenting Author
1	Local adaptation at the transcriptome level in Daphnia populations	Ravindran, Ms. Suda Parimala Universität Hamburg
5	Genetic covariance between additive and dominant effects in closed lines selected for more than 40 generations	Legarra, Andres INRA
8	Effect of genetic architecture and sample size on the accuracy of genomic prediction of complex traits	Morgante, Fabio North Carolina State University
11	The use of marker-data transformations to account for linkage disequilibrium in genomic selection: a case study in switchgrass (Panicum virgatum L.)	Ramstein, Mr. Guillaume University of Wisconsin-Madison
14	Inbreeding coefficient under a grandparental regression model	Munilla, Dr. Sebastián University of Buenos Aires
17	Bagging elastic net	Xavier, Alencar Purdue University
20	Holistic selection for eggshell quality using a bio-complex in layer chicken	Arango, Dr. Jesus Hy-line International
25	Does selection act on gene networks? A case study in the model plant Medicago truncatula.	Bonhomme, Mr. Maxime Université Toulouse III
28	Bayesian-information and Linkage-disequilibrium iteratively nested keyway (BLINK) model to solve both power and computing problems in genome-wide association studies	Huang, Dr. Meng Washington State University
31	The application of novel phenotyping technologies to reveal the genetic basis of dynamic stress-responsive traits in cotton	Pauli, Dr. Duke Cornell University
34	Bioeconomic selection index (EMGTe) for Nellore Brazil breeding program	Lobo, Dr. Raysildo USP
40	Rare variants from imputed whole genome sequence explain a small fraction of the total genetic variance in different traits in cattle	Zhang, Mrs. Qianqian Aarhus University
45	Genome-wide association analysis in dogs implicates 98 loci as risk variants for cranial cruciate ligament rupture	Baker , Lauren University of Wisconsin-Madison
48	Sequence-level partitioning of genomic variance for mastitis and production trains in dairy cattle based on transcrptomic expression data	Fang, Mr. Lingzhao Aarhus University
51	Genomic prediction accuracies in commercial barley and wheat schemes	Edriss, Dr. Vahid Nordic Seed
54	Polymorphisms in MCT1 and CD147 genes in Arabian and Quarter Horses	Queiroz-Neto, Dr. Antonio FCAV-UNESP
57	Genome-wide association study for reproduction traits in Holstein cattle	Guarini, Ms. Aline University of Guelph
61	Integrations of risk factors, copy number variants, and gene expression to predict survival of breast cancer patients	Gonzalez-Reymundez, Agustin Michigan State University

64	The effects of alternative selective regimes on the genetic basis of adaptation to novel condition in experimental populations	Huang, Dr. Yuheng University of Wisconsin-Madison
67	Genetic evidence of mating for height and body mass index in humans	Robinson, Dr. Matthew University of Queensland
72	Which individuals to phenotype? Optimal design of reference population for genomic selection while maintaining genetic diversity	Eynard, Ms. Sonia INRA / Wageningen UR
75	Factors affecting the estimation of the genetic correlation between populations	Wientjes, Ms. Yvonne Wageningen UR
78	Phenotypic effect of recessive embryonic lethal haplotypes on non-return rate in cattle	Wu, Dr. Xiaoping QGG, Aarhus University
81	Multiple copies of alkane metabolic genes in bacteria	Masuda, Dr. Hisako Indiana University Kokomo
84	When Bayesians encounter Amdahl's law - A statistical perspective of parallel computing for genomic prediction	Wu, Dr. Xiao-lin Geneseek
88	Modelling the relationship between social interactions and inherited variability	Marjanovic, Ms. Jovana Wageningen University/SLU
92	Controlling population structure in the genomic prediction of tropical maize inbreds	Lyra, Ms. Danilo Iowa State University
96	Additive and non-additive genetic effects for disease resistance in the Pacific White Shrimp Penaeus vannamei	Campos-Montes, Dr. Gabriel Ricardo Universidad Autónoma Metropolitana
99	A novel pedigree-free linkage analysis using multiple regression	Mei, Dr. Bujun Hetao College
102	Candidate variants confirmation by GWAS applied to large cow data sets	Marete, Mr. Andrew INRA
105	Leveraging the use of genomic resources to guide genomic selection	Pino Del Carpio, Dr. Dunia Cornell University
108	Recombination and genetic variance among maize doubled haploids induced from F1 and F2 plants	Sleper, Joshua Syngenta
113	Bayesian inference of the allelic series at quantitative trait loci in multiparent populations	Crouse, Wesley University of North Carolina-Chapel Hill
114	Genomic data integration for GWAS and eQTL analysis	Rojo, Constanza University of Wisconsin-Madison
117	Contemporary group estimates with the inclusion of minimum monthly temperature in the model increases the ability to detect genotype by environment interactions in pigs	Guy, Mrs. Sarita The University of Sydney
120	Using metafounders reduces bias in single step GBLUP genomic evaluations	Vitezica, Dr. Zulma INRA
123	Distinct genetic architectures for maize phenotypes and phenotypic plasticity	Kusmec, Aaron Iowa State University
127	A simple test for genome-wide evidence of selection on complex traits	Beissinger, Dr. Timothy USDA-ARS/University of Missouri
132	Implications of genetic variability using bulk within progenies derived from F2 plants in common bean	Andrade Pereira, Lais Universidade Federal De Lavras/ Purdue

135	Associations with flowering time, latitude, and climate in switchgrass	Grabowski, Dr. Paul USDA-ARS
139	Evaluation of efficiency of marker-assisted seedling selection in an apple breeding program	Ru, Mrs. Sushan Washington State University
142	Genetic parameters for beef tenderness over various post-mortem aging times determined with a random regression model	Miller, Dr. Stephen AgResearch
146	Multi-breed genomic prediction using single- and multi-trait models	Calus, Dr. Mario Wageningen UR
149	Genomic prediction of early stage single-cross hybrids in maize	Kadam, Mr. Dnyaneshwar University of Minnesota-Twin Cities
152	Contribution of social (or indirect) genetic effects to health and disease - a study in laboratory mice.	Baud, Dr. Amelie European Bioinformatics Institute
155	Genetics of skeletal evolution in unusually large mice from Gough Island	Parmenter, Michelle University of Wisconsin-Madison
158	Prediction of agronomic traits in maize using new source of data	Xu, Yang University of California-Riverside
161	Development and application of a bioinformatics pipeline for genotyping-by-sequencing (GBS) of autotetraploid potato.	Smith, Schuyler University of Wisconsin-Madison
164	Identifying deleterious mutations in maize	Lu, Dr. Fei Cornell University
167	Virgin Genomics: opportunities for emerging sectors - SNP chips optional	Rowe, Dr. Suzanne Agresearch Limited
170	Divergence and admixture in Chinese and European Sus scrofa	Guldbrandtsen, Dr. Bernt Aarhus University
173	Translating genomics into competitive products for the market	da Silva, Dr. Sofia KWS
176	Use of individual fitness in analyzing animal lifetime performance	Maki-Tanila, Dr. Asko University of Helsinki
179	Genome prediction for MIR predicted groups of milk fatty acids in Holstein cattle	Cruz, Dr. Valdecy UoG/CNPq
182	Simultaneous modeling of phenotype mean and variance: Rationale, results, and software tools	Corty, Robert University of North Carolina-Chapel Hill
185	Rare variants and parent-of-origin effects on whole blood gene expression identified in large family pedigrees	Viñuela , Dr. Ana University of Geneva
189	On the trait-specific correlation matrix of genomic breeding values	Teuscher, Dr. Friedrich Fbn Dummerstorf
192	Effect of shrinkage on prediction accuracy of genomic selection in tomato population	Duangjit, Dr. Janejira Kasetsart University
195	Genome wide association analysis of genetic variants of virus diseases in Cassava	Kayondo, Mr. Siraj Ismail Cornell university
198	Characterization of genetic diversity, population structure and linkage disequilibrium within and across 35 European maize landraces using high-density genomic data	Mayer, Manfred Technical University of Munich
201	Corn hybrids stability and adaptability by Bayesian-AMMI statistical procedures for cercosporiose	Milani Craft, Dr. Valentina Universidade Estadual De Maringá

204	Applying genomic selection in beef cattle: a deterministic simulation	Gordo, Dr. Daniel UNESP/Jaboticabal
209	The inclusion of transcriptome information in genome-wide association studies and genomic selection models to improve the prediction ability	Sant'Anna, Ms. Isabela University of Florida
211	Exploring properties of genome wide association analyses based on broadly different prior specifications	Chen, Chunyu Michigan State University
213	Developing a method to screen for salt tolerance in carrot	Bolton, Mr. Adam University of Wisconsin-Madison
216	Prospects of genomic prediction for Goss's bacterial wilt and leaf blight resistance using bi-parental populations and a diversity panel of Maize	Singh, Amritpal University of Minnesota
219	Genomic regions involved in seed protein, oil, and carbohydrate concentrations in glycine max	Mcconaughy, Ms. Samantha University of Nebraska-Lincoln
222	Investigating the genetic architecture of complex traits in a barley NAM population	Ollhoff, Alex University of Minnesota
225	Patterns of differentiation between house mouse subspecies at loci implicated in hybrid sterility: Do loci contributing to reproductive isolation show distinctive signatures in the genomic landscape?	Turner, Dr. Leslie Max Planck Inst. For Evolutionary Biology
228	The effect of PRRS virus disease outbreak on genetic parameters of reproductive performance in pigs	Putz, Mr. Austin Iowa State University
232	Construction of single-sample gene networks using transcriptomic and functional information for network QTL mapping	Dong, Ms. Qian (Jessie) Iowa State University
236	Alternative approaches to modeling breeding value in tree breeding programs	Festa, Adam North Carolina State University
241	The genetic basis of the coordination between resource allocation and resource acquisition in Drosophila melanogaster	King, Dr. Elizabeth University of Missouri
244	Use of environmental convariates for genomic selection modeling genotype by environment interaction in a wheat bredding program	González Barrios, Mr. Pablo Martín Universidad De La República
247	Applying quantitative genetics advances towards molecular breeding in rice.	Thomson, Dr. Michael Texas A&M University
250	Trans regulatory architecture of genetic transcriptome variation from 1,000 yeast individuals	Albert, Dr. Frank University of Minnesota
253	Modeling genotype by environment interaction over multiple covariates for feed efficiency in dairy cattle	Tempelman, Dr. Robert Michigan State University
256	Joint local testing of variant-sets improves power and interpretation of gene-by-context interactions	Casale, Mr. Francesco European Bioinformatics Institute
262	A hyperparameterized whole genome simulator for developing and evaluating genetic designs for population and quantitative genetic studies	Dougherty, John University of Delaware
266	Fine-mapping a major maize domestication QTL for ear diameter	York, Mrs. Alessandra University Of Wisconsin-Madison

269	Genetic Architecture of the Maize Inflorescence Shattering Trait	Tuholski, Michael University of Wisconsin-Madison
272	GWAS for Resistance to Stem Rot and Aggregated Sheath Spot of Rice	Rosas, Juan INIA
275	Dissecting the Genetic Architecture of Local Adaptation Using Environmental Association and Selection Mapping in Soybean	Bandillo, Mr. Nonoy University of Nebraska-Lincoln
278	Optimizing the Training Population Composition to Improve Genomic Prediction Accuracy across Selection Cycles of a Barley Breeding Population	Tiede, Tyler University of Minnesota

TUESDAY POSTERS

Poster Number	Title	Presenting Author
3	Spontaneous mutations and the origin and maintenance of quantitative genetic variation	Huang, Wen North Carolina State University
6	Accounting for randomness of genotypes in across and single population genome-enabled prediction: A hierarchical Bayes approach	Martinez, Mr. Carlos University of Florida
9	Functional validation of loci contributing to nicotine resistance in Drosophila	Highfill, Mr. Chad University of Kansas
12	Quantitative exploration of Ethiopian durum wheat: joining genetics and social sciences to produce a better wheat through GWAS and a multiparental population design	Dell'Acqua, Dr. Matteo Scuola Superiore Sant'Anna
15	The genetic basis of evolved acute stress response in the nematode Caenorhabditis remanei	O'Connor, Ms. Christine University of Oregon
18	When one plus one does not equal two: some tandem gene duplicates are overactive	Loehlin, Dr. David University of Wisconsin-Madison
21	Genome-wide and species-wide variation in C. elegans reveals association of telomere length with population differences in pot-2	Cook, Daniel Northwestern University
22	Natural variation alters sensitivities to topoisomerase II poisons through changes in conserved drug-binding residues	Zdraljevic, Stefan Northwestern University
26	Covariance association test (CVAT) identify genetic markers associated with schizophrenia in functionally associated biological processes	Rohde, Palle Aarhus University
29	Evaluation of merit of including non-additive genetic variation in genomic prediction of feed-related traits in broiler chickens	Li, Dr. Yutao CSIRO Agriculture
32	Studying temperature and salinity stresses in the model brown alga Ectocarpus sp. by QTL mapping	Avia, Dr. Komlan CNRS-Station Biologique Roscoff
35	Exploring the last chromosome: Y-linked sequence variation in the house mouse	Morgan, Mr. Andrew University of North Carolina
38	Heritability of mentality traits in Swedish dogs	Strandberg, Dr. Erling Swedish University of Agricultural Sciences

41	Joint genetic evaluation of mastitis resistance and recovery ability in danish dairy cows under automated milking systems	Welderufael, Mr. Berihu Swedish University of Agricultural Sciences
46	Community genomics of plant-insect interactions	Barker, Mrs. Hilary University of Wisconsin-Madison
49	Prediction for candidates two or more generations away from the reference population	Ma, Dr. Peipei Aarhus University
52	Mapping variants to gene ontology categories improves genomic prediction for quantitative traits in Drosophila melanogaster	Sørensen, Dr. Peter Aarhus University
55	Accuracies of genomic prediction of breeding values for growth, carcass and meat quality traits in a crossbred sheep population including non-additive effects	Brito, Luiz University of Guelph
58	Diallel analysis of top size in carrot (Daucus carota, L.) using biplots and image-based phenotyping	Turner, Ms. Sarah University of Wisconsin-Madison
62	Comparison between haplotype-based and single SNP-based methods for genomic predictions in Holstein cattle	Karimi, Mrs. Seyedehzahra University of Guelph
65	Insights into the genetic architecture of bioenergy traits in eastern cottonwood from a genome-wide association study including common and rare genetic variants	Fahrenkrog, Annette University of Florida
69	Prediction of growth curves in pigs using a hierarchical Bayesian model	Duijvesteijn, Dr. Naomi Topigs Norsvin Research Center
70	Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations	Brinker, Mrs. Tessa Wageningen University
73	Candidate gene identification for maize gibberella ear rot disease resistance using genomic and transcriptomic approaches	Kebede, Dr. Aida Agriculture and Agri-Food Canada
76	Consequences of model misspecification on REML estimates of heritability using genomic models	Cuyabano, Dr. Beatriz Aarhus University
79	A whole genome association study of yearling weight, carcass traits and primal-cuts of Hanwoo (Korean brown cattle)	Choi, Dr. Tae Jeong National Institute of Animal Science
82	Data structure requirements for estimation genotype-environment interactions when genomic information is available	Fikse, Dr. Freddy SLU
86	SNP discovery in wheat RIL population using genotyping-by-sequencing and genome-wide QTL mapping for plant height	Hussain, Mr. Waseem University of Nebraska-Lincoln
89	Tunable combinatorial transcriptional engineering based on CRISPR interference	Tarasava, Ms. Katia University of Colorado-Boulder
93	Patterns of genomic and phenomic diversity in grape and apple	Migicovsky, Zoe Dalhousie University
97	The genetic architecture of gene expression in human peripheral blood	Lloyd-Jones, Dr. Luke Queensland Brain Institute
100	Identification of genomic islands from sequence data in Atlantic cod (Gadus morhua L.)	Rodriguez Ramilo, Dr. Silvia Teresa INRA
103	Inclusion of non-additive effects on G-BLUP model considering different kinship matrices	Vieira, Mr. Indalécio UFLA

106	Pigmentation and pupation height phenotypic plasticity catalyzed by desiccation in worldwide populations of Drosophila melanogaster	Simard, Ms. Audrey University of Wisconsin-Madison
110	Whole genome hitchhiking on an organelle mutation	Flood, Dr. Padraic Max Planck Inst for Plant Breeding Res.
115	Multivariate genome-wide association study (GWAS) for detailed milk protein composition	Teklewold, Mr. Grum Aarhus University
118	Identifying recombination events and haplotypes in beef cattle from half-sib families	Ferdosi, Mohammad University of New England
121	Genetic analysis of productive life in Chinese Holstein population	Wang, Yachun China Agriculture University
125	Recombination patterns among spring barley populations	Poets, Dr. Ana University of Minnesota
128	Artificial selection for odor-guided behavior in Drosophila melanogaster	Brown, Ms. Elizabeth University of Cincinnati
133	Association study of lactation performance in a dairy sheep population using classic GWAS and genomic models	Li, Hao University of Wisconsin-Madison
136	Common SNPs explain a greater proportion of complex trait heritability than previously thought	Speed, Doug University College London
140	Signatures of selection in sweet maize suggest a single donor shapes modern elite se1 inbreds	Murray, Matthew University of Wisconsin-Madison
144	GWAS of tetraploid potato with automated genotype calls	Schmitz Carley, Cari University of Wisconsin-Madison
147	The genetics of just right: Dose-response surface fits to drought and nitrogen limitation applied together allow mapping of maize loci that exhibit nonlinear responses	Stapleton, Ann University of North Carolina- Wilmington
150	Genetic parameter estimation and genome wide association study for NDV response in chickens	Rowland, Kaylee Iowa State University
153	Genomic regions associated with feed efficiency and component traits of pigs fed lower energy, higher fiber diets after divergent selection for residual feed intake	Mauch, Ms. Emily Iowa State University
156	The effect of crossing strategy on genomic prediction in maize	Burdo, Brett University of Wisconsin-Madison
159	Incorporating transcriptomics into genomic selection to predict quantitative resistance to Cassava Brown Streak Disease (CBSD)	Gonzalez del Valle, Mr. Roberto Lozano Cornell University
162	Multivariate index improves genomic selection accuracy and gain for disease resistance in maize	Willmot, Dr. David AgReliant Genetics
165	Genome-wide association study of seed protein and oil content in soybean nested association mapping population	Salari, Mr. Mohammad Purdue University
168	The value of long term multi-year multi-location trial data to evaluate risk in prediction of early cultivar performance	Arief, Dr. Vivi University of Queensland
171	Nonparametric genome-wide prediction and variable interaction with Bayesian additive regression trees	Waldmann, Dr. Patrik Swedish University of Agricultural Sciences

174	Development of a phenotypic platform for evaluation of kernel processing quality for Frito Lay corn chip production	Holmes, Mr. Mark University of Minnesota
177	Genomic prediction in Coffea canephora using Bayesian polygenic modeling	Ferrão, Mr. Luis Felipe Felipe
180	Investigating the genetics of selection response for flowering time in a multi-environment parallel selection experiment	Manching, Mrs. Heather University of Delaware
183	Genome-wide family prediction	Munoz, Dr. Patricio University of Florida
186	Predicting causal variants affecting expression using whole genome sequence and RNA-seq from multiple human tissues	Brown, Dr. Andrew University of Geneva
187	Genome-wide association study of milk quality traits in Brazilian water buffaloes using phenotypes from non-genotyped animals	Barros, Ms. Camila São Paulo State University
190	The influence of organic and conventional production systems on breeding for carrot top height	Grahn, Charlene University of Wisconsin-Madison
193	Bayesian AMMI analysis of genotype-by-environment interaction in common beans in Paraná- Brazil	Fonseca, Dr. Ines University of Wisconsin-Madison
196	Genomic selection of an index trait to develop winter malting barley	Falcon, Celeste University of Wisconsin
199	Estimation of effective population size from linkage disequilibrium measures in the Spanish Holstein population	Fernández, Jesús INIA
202	Partitioning variance in maize populations based on genome annotations	Bradbury, Peter USDA-ARS
205	Quantitative genetic analysis of major secreted glycoproteins in a mouse model of asthma	Donoghue, Lauren University of North Carolina-Chapel Hill
210	An expanded Wisconsin maize diversity panel and its application in GWAS of flowering time	Mazaheri, Mona University of Wisconsin-Madison
214	Maintaining the accuracy of genomewide predictions when selection has occurred in the training population	Brandariz Zerboni, Sofia University of Minnesota
217	Construction of high density linkage maps and detection of downy mildew resistanceLocus in a Vitis aestivalis-derived 'Norton' population	Sapkota, Mr. Surya University of Missouri & MSU
220	Construction of genetic maps in complex autopolyploids	Mollinari, Dr. Marcelo North Carolina State University
223	Genotype calling and linkage mapping from GBS data in Ipomoea trifida, a highly heterozygous species	Da Silva Pereira, Guilherme North Carolina State University
226	Alternative models for genetic analysis of tick and worm counts in Nellore cattle	Passafaro, Tiago Luciano University of Wisconsin-Madison
230	GxE models for performing predictions in complex scenarios: i) Tested genotypes in untested environments, and ii) Untested genotypes in untested environments	Jarquin, Juan Diego University of Nebraska-Lincoln
233	A shifting genetic architecture underlies selection response for flowering time adaptation in maize	Wisser, Dr. Randy University of Delaware
234	Characterizing the allele-frequency spectrum of causal variants in yeast	Bloom, Dr. Joshua University of California- Los Angeles

239	Identification of QTLs involved in quinolizidine and indole alkaloids content in Lupinus luteus L.	Osorio, Dr. Claudia CGNA
242	Studying the genetics of microRNA expression for alcohol related traits [in a panel of recombinant inbred mouse strains]	Rudra, Dr. Pratyaydipta University of Colorado
245	Post-hoc blocking and genotype by environment interaction in Zossiagrass	Xing, Lin University of Florida
248	Hybrid prediction effectiveness using genome-wide information in tobacco	Carvalho, Dr. Bruna British American Tobacco
251	Functional SNP in a polygenic disease induced by high-altitude in fattening Angus steers by combining multi-OMICS data into Systems Biology	Canovas, Dr. Angela University of Guelph
254	Working on Evolvix, the first general-purpose programming language designed by biologist for biologists.	Loewe, Dr. Laurence University of Wisconsin-Madison
259	Phenotypic variation and genetic dissection of silage yield and compositional traits in recombinant inbred testcrosses in Maize. (Zea mays L.)	Renk, Jonathan University of Wisconsin-Madison
263	Statistical methods for the functional characterization of selection signatures	Servin, Dr. Bertrand INRA
267	Characterization of the genetic architecture of maize ancestor, Teosinte	Yang, Chin Jian University of Wisconsin-Madison
270	Enhancement of Genomic Prediction and QTL-Mapping in Bovine Respiratory Disease using Sequence Imputation and Feature Selection	Hoff, Jesse University of Missouri
273	Fine Mapping Coincident QTL for Multiple Traits on 6H of Barley - The Use of Traditional Method to Resolve Unfavorable Associations of Quantitative Traits	Yin, Lu University of Minnesota
276	A widely-applicable and powerful method for detecting balancing selection	Siewert, Ms. Katherine University of Pennsylvania
279	Fine-mapping of QTL Associated With Genotype-Dependent Plant Regeneration in Maize	McFarland, Mr. Frank University of Wisconsin-Madison

THURSDAY POSTERS

Poster Number	Title	Presenting Author
4	Higher male than female recombination rate in cattle is controlled mostly by genetic variants effective in both sexes	Druet, Tom Unit of Animal Genomics (ULg)
7	Dissecting the genetic and environmental contributions to anthropometric and cardio-metabolic trait variation in humans.	Xia, Charley University of Edinburgh
10	GGE biplot of genotype by environment interaction for tuber yield of potato clones resistant to potato virus Y	Andrade, Ms. Mario Universidade Federal de Lavras
13	Asymptotic normality of breeding values under segmental inheritance.	Cantet, Dr. Rodolfo University of Buenos Aires-Inpa Conicet

16	Genome-wide association studies of additive and dominance effects for female fertility traits in Danish Holstein cattle	Mao, Mr. Xiaowei Aarhus Univ./Swedish Univ. of Ag. Sci.
19	Prediction of top cross performance in maize for drought tolerance via genome-wide markers	Silva, Dr. Luciano SAS Institute Inc.
23	Genetic control of fumonisin incidences in tropical maize	Souza/FUNDECC, Dr. Joao Federal University of Lavras
27	Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction	Zhou, Mr. Yao Washington State University
30	The estimation of additive, dominance and epistatic effects underlying lameness in Fleckvieh and Braunvieh cows	Suchocki, Dr. Tomasz Wroclaw Univ. of Environmental & Life Sci.
33	Accuracy of genomic breeding values for meat tenderness in Polled Nellore cattle	Magnabosco, Dr. Claudio EMBRAPA
36	Modeling genome by environment interaction in allohexaploid wheat (Triticum aestivum)	Santantonio, Nicholas Cornell University
39	Genetic mapping by bulk segregant analysis in Drosphila: Experimental design and simulation-based inference	Pool, John University of Wisconsin-Madison
44	Proportion of inter-individual differences for time to death of glioblastoma multiforme incorporating multi-omic layers.	Bernal Rubio, Yeni Liliana Michigan State University
47	Bayesian networks illustrate phenotypic and genomic trait connections in maize	Toepner, Ms. Katrin Technical University of Munich
50	Assessment of bias due to genomic pre-selection in the Dutch-Flemish breeding value estimation	Eding, Dr. Herwin Crv
53	Co-expression gene networks reveal potential biomarkers for reduction of boar taint in pigs	Kadarmideen, Professor Haja University of Copenhagen
56	Bias due to selective genotyping in genomic prediction using H-BLUP	Wang, Ms. Lei Aarhus University
60	Trait-specific long-term consequences of genomic selection in beef cattle	Queiroz, Dr. Sandra FCAV - UNESP
63	Prospects for genomic selection in macadamia, an outcrossing perennial rainforest tree	O'Connor, Katie University of Queensland
66	On curious properties of genomic relationship matrices in mixed models	Meyer, Prof. Karin University of New England
71	Bias of estimated allele substitution effects due to dominance, for two models of gene-action	Duenk, Mr. Pascal Wageningen University
74	Within-family genomic selection in aquaculture breeding programmes	Toro, Dr. Miguel Universidad Politécnica De Madrid
77	A double repeatability model to evaluate shell quality in layer chicken	Wolc, Dr. Anna Iowa State University
80	Chemotherapy-induced neutropenia is modulated by drug-specific genes	Gatti, Daniel The Jackson Laboratory
83	Including crossbreds in the genomic relationship matrix through utilization of both linkage analysis and linkage disequilibrium	Iversen, Ms. Maja Topigs Norsvin
85	Use of electronic health record to predict family relationships for genetic epidemiology research	Huang, Mr. Xiayuan University of Wisconsin-Madison

87	Optimizing size and sex ratio of the genotyped population for genomic selection in broilers using information from both genotyped and non-genotyped individuals	Alemu, Dr. Setegn Aarhus University
90	Genomic prediction in cattle based on sequence data	Schuler, Urs Qualitas AG
94	The identification of haplotypes associated with major depressive disorder	Howard, Dr. David University of Edinburgh
95	The epistasis boundary: Linear vs. nonlinear genotype-phenotype relationships	Sverdlov, Dr. Serge University of Washington
98	QTL mapping and genomic selection methods for assisted breeding in the American cranberry	Covarrubias, Giovanny University of Wisconsin-Madison
101	Identification of the causal mutation for transverse hemimelia in River Buffalo using whole genome sequence data	Whitacre, Lynsey University of Missouri
104	Marker-based estimates reveal significant non-additive effects in clonally propagated cassava (Manihot esculenta): implications for the prediction of total genetic value and the selection of varieties	Wolfe, Dr. Marnin Cornell University
107	Rare variants explain a substantial fraction of the genetic variance for female fertility in dairy cattle	Sahana, Dr. Goutam Aarhus University
109	Genomic prediction based on interaction networks of markers: How to incorporate prior experimental information	Martini, Dr. Johannes Georg-August University
112	Epigenome-wide association study and prediction of triglyceride postprandial responses to high-fat dietary challenge	Lai, Dr. Chao-Qiang USDA ARS HNRCA At Tufts University
116	VALIDATE on the CyVerse cyberinfrastructure: scalable testing of the accuracy and precision of association and prediction software	Chaung, Danielle University of North Carolina-Wilmington
119	Linkage drag connects spike and root development in wheat	Voss-Fels, Mr. Kai Justus Liebig University Giessen
122	Genetic parameter estimation in purebred and crossbred pigs including dominance with kernel and two genomic models	Tusell Palomero, Dr. Llibertat INRA
126	Polymorphisms in PDK4 and DMRT3 genes in Arabian and Quarter Horses	Regatieri, Inaê UNESP-FCAV Jaboticabal
130	High density genome scan for selection signatures in French sheep reveals allelic heterogeneity and introgression at adaptive loci	Rochus, Christina Swedish University of Agricultural Sciences
134	Developing a framework to leverage soybean uniform regional test data for genomic prediction modeling	Nice, Liana University of Minnesota
137	Parallel evolution of ethanol tolerance found in four populations of Drosophila melanogaster	Sprengelmeyer, Quentin University of Wisconsin-Madison
141	Genetic correlation between gain and carcass traits in purebred and crossbred pigs	Godinho, Mr. Rodrigo Wageningen Univ./Univ. Federal de Viçosa
145	Meat quality traits genome-wide association study in Nellore cattle	Albuquerque, Dr. Lucia UNESP

148	Investigating causal relationships underlying phenotypic traits of two fruit species of the Sapotaceae family via graphical models	Pinto, Mr. Renan University of Wisconsin-Madison
151	Genomic genetic evaluation in young animals using different methods	Tonussi, Mr. Rafael Rafael
154	Genomic prediction within and between subpopulations of the USA national maize inbred collection	Gage, Joseph University of Wisconsin-Madison
157	Evolution of gene expression in giant mice from Gough Island	Nolte, Mark University of Wisconsin-Madison
160	Predicting genetic variance in barley using Genotyping-by-Sequencing (GBS) Markers and Population Sequencing (POPSEQ)	Neyhart, Jeff University of Minnesota
163	The impact of rare alleles on expression and fitness in maize	Kremling, Mr. Karl Cornell University
169	Reparametrization-based estimation of genetic parameters in multi-trait animal model using Integrated Nested Laplace Approximation	Mathew, Dr. Boby University of Bonn
172	Identifying genetic risk factors for Drug-Induced Liver Injury: Detection of susceptible strains and potential vulnerability QTL in the Collaborative Cross mouse population	Kim, Dr. Yunjung University of North Carolina-Chapel Hill
175	Genetics of hybrid performance in maize: QTL detection for biomass production in a reciprocal multiparental design	Charcosset, Dr. Alain INRA
178	Modeling the variance-covariance matrix of genetic and residual effects in a Panicum maximum genome wide selection experiment	De Castro Lara, Ms. Letícia Letícia Lara
181	Genome-based prediction of testcross performance with different testers in hybrid rye breeding	Lehermeier, Dr. Christina Technical University of Munich
184	Accounting for genetic effects as confounders in propensity score analysis	Cardoso Ferreira, Mrs. Vera University of Wisconsin-Madison
188	Genetically and agronomically characterizing a core-set of interspecific breeding lines	Eickholt, Mr. David North Carolina State University
191	Improving NGS-SNP identification and inference in tetraploids; a white clover example	Bhakta, Dr. Mehul University of Florida
194	Improving genomic prediction by adjusting for spatial correlation and genetic competition in cassava field experiments	Elias, Dr. Ani Cornell University
197	Efficient QTL mapping of additive and dominance effects in outbred multiparent populations	Keele, Mr. Gregory University of North Carolina-Chapel Hill
200	A statistical framework for leveraging sequence data from supposed technical replicates of an individual to detect experimental errors	Chan, Ariel Cornell University
203	Drone-based high-throughput phenotyping for the selection of high-yielding soybean lines	Moreira, Ms. Fabiana Purdue University
206	Investigation of genotype by environment barriers to maize adaptation using near-isogenic lines capturing an allelic series	Wills, Dr. David USDA-ARS

207	Genetic analysis of an intermediate phenotype for recombination rate variation	Wang, Richard University of Wisconsin-Madison
212	Genome-based analysis of stayability in Nellore cattle using the single-step genomic BLUP approach	Fernandes Jr., Dr. Gerardo Sao Paulo State University - FCAV/UNESP
215	Genomic selection shows promise for improving winter wheat: Insights from the University of Nebraska-Lincoln wheat breeding program	Belamkar, Dr. Vikas University of Nebraska-Lincoln
218	Genome-wide association analyses based on alternative strategies for modeling feed efficiency	Lu, Ms. Yongfang Michigan State University
221	Accuracy of genomewide prediction with and without heterozygote effects in apple	Blissett, Elizabeth University of Minnesota
224	Performance of fixed vs. variable length haplotypes for genomic prediction	Hayr, Melanie Iowa State University
227	Combining marker and pedigree information may enhance multiple-trait genome-enabled prediction	Brito Lopes, Dr. Fernando University of Wisconsin-Madison
231	Modeling temporal estimates of genotype x environment interaction for Nile tilapia in three production systems	Fernandes, Arthur University of Wisconsin-Madison
235	Multiple-trait genomic selection methods to increase prediction accuracy in wheat quantity	Lado, Mrs. Bettina Universidad De La República
240	Accuracy of estimated breeding values for hoof lesions in Canadian Holsteins	Malchiodi, Francesca CGIL
243	Comparing genotyping methods for development of genomic selection prediction models in Eucalyptus	Furtado dos Santos, Rodrigo University of Florida
246	Super panel development and genomic analyzes using two high-density SNP platforms in Nellore beef cattle	Ventura, Dr. Ricardo University of Guelph
249	Performance of genomic prediction for a sugarcane commercial breeding program	Brum, Itaraju Cornell University
252	Impact of relatedness on genomic prediction and GWAS detection in two elite eucalyptus breeding populations	Müller, Ms. Bárbara University of Brasília
255	Investigating kinship and heterotic relationships of expired maize plant variety protection lines	White, Mr. Mike University of Wisconsin-Madison
260	Breeding strategies and multivariate approaches to optimizedrought resistance in plants	Bodah, Dr. Eliane University of Washington
264	Genomics-aided Development of Cucumis hystrix Introgression Library for Cucumber Improvement	Thammapichai, Paradee University of Wisconsin-Madison
268	Heritability and genetic gain of Bipolaris resistance in switchgrass	Songsomboon, Mr. Kittikun Cornell University
271	Biochemical Analysis of Epicuticular Wax in Onions: a Thrips Resistance Strategy.	D. Munaiz, Eduardo University of Wisconsin-Madison
274	Shared selective sweeps across human populations	Johnson, Kelsey University of Pennsylvania
277	Vernalization, dormancy, and the annualization of onion (Allium cepa) for breeding	D'Angelo, Chris University of Wisconsin-Madison
282	The construction and utility of a high-resolution meiotic crossover map in potato	Marand, Alexandre University of Wisconsin-Madison
295	Genome-wide association study for female fertility traits in Chinese holstein population	Liu, Aoxing Aarhus University

POSTER ABSTRACTS

1

Local adaptation at the transcriptome level in Daphnia populations

Miss Suda Parimala Ravindran[1], Prof.Dr. Mathilde Cordellier[1]
[1]*Universität Hamburg*

Local adaptation and its underlying molecular basis has long been the focus of evolutionary biology studies. Recently there has been increased interest in the evolutionary role of Daphnia's plasticity and the molecular mechanisms underlying local adaptation. Daphnia plays a central role in aquatic ecosystems, has a well-known ecology and is widely used in population studies and toxicology. The aim of the present study was to identify candidate genes for local adaptation in Daphnia, understand their role and response to environmental stressors, and analyze the sequence polymorphism. Using transcriptome analysis, we assessed differences in gene expression profiles within- and among- four Daphnia galeata populations in Europe. In total, about 4% of the 32903 genes were differentially expressed between the populations. To understand whether the observed difference in expression is merely due to genetic drift or rather selection, we analyzed the distribution of Q_{ST} values. Although Daphnia's ecology is intensively studied, little is known on the functional underpinning of genotypic responses to environmental stressors. Using sequence information, the transcripts were annotated using a comparative genomics approach. The extensive body of literature available for Daphnia species allowed identifying stressors linked with the observed gene expression patterns. Furthermore, a SNP analysis allowed inferring population structure and the distribution of genetic variation. The population divergence at the sequence level was higher than the gene expression divergence by several orders of magnitude. Further analysis will allow defining the respective role of strong founder effects, genetic drift and selection in the observed pattern. This study allows us to understand the genetic background of adaptation to environmental changes in a key species of aquatic ecosystems.
Keywords: Daphnia galeata – gene expression - Q_{ST} - adaptation – SNP - functional annotation - stressor

3

Spontaneous mutations and the origin and maintenance of quantitative genetic variation

Wen Huang[1], Richard F. Lyman[1], Rachel Lyman[1], Mary Anna Carbone[1], Susan Harbison[1], Michael Magwire[1], Trudy F.C. Mackay[1]
[1]*North Carolina State University*

Mutation and natural selection together shape the genetic variation in natural populations; natural selection of different forms can either eliminate or preserve genetic variation generated by mutations. Here, we compare genetic variation in Drosophila melanogaster mutation accumulation (MA) lines maintained with minimal natural selection to genetic variation among wild-derived inbred lines, which results from mutation, genetic drift, and natural selection, thus enabling us to investigate the origin and maintenance of genetic variation. The median spontaneous mutation rate was estimated to be 6.60e-9 events per base, with considerable variation among the MA lines. On average, spontaneous mutations generated quantitative genetic variation at a rate of 0.65e-3 Ve per generation for gene expression traits and 1.36e-3 Ve for organismal phenotypes, where Ve is the environmental variance. The magnitude of genetic variation among the wild-derived inbred lines was much lower than predicted from a neutral model of spontaneous mutation-genetic drift balance, and the mutational effects were much larger than allelic effects of standing polymorphisms, suggesting that strong stabilizing selection constrains genetic variation for quantitative traits in nature. However, simple models of mutation-stabilizing selection balance cannot account for strong stabilizing and the observed estimates of mutational and standing genetic variance, motivating further empirical work to assess the balance of evolutionary forces responsible for quantitative genetic variation.

4

Higher male than female recombination rate in cattle is controlled mostly by genetic variants effective in both sexes

Naveen K Kadri[1], Chad Harland[1,2], Pierre Faux[1], Nadine Cambisano[1,3], Latifa Karim[1,3], Wouter Coppieters[1,3], Sébastien Fritz[4,5], Erik Mullaart[6], Didier Boichard[5], Richard Spelman[2], Carole Charlier[1], Michel Georges[1], Tom Druet[1]

[1]Unit of Animal Genomics (ULg), [2]Livestock Improvement Corporation, [3]Genomics platform, GIGA, [4]Allice, [5]GABI, INRA, [6]CRV

We herein study genetic recombination in three dairy cattle populations from France, New-Zealand and the Netherlands. We identify 2,395,177 crossover (CO) events in sperm cells transmitted by 2,940 sires to 94,516 offspring, and 579,996 CO events in oocytes transmitted by 11,461 cows to 25,332 offspring. When measured in identical family structures, the average number of CO in males (23.3) was found to be larger than in females (21.4). The heritability of global recombination rate (GRR) was estimated at 0.13 in males and 0.08 in females. The genetic correlation was equal to 0.66, indicating that shared variants are influencing GRR in both genders. Haplotype-based genome-wide association studies revealed seven genome-wide significant QTL. Variants identified by next-generating sequencing in 5 Mb windows encompassing the QTL peaks were imputed in order to perform a sequence-based association analysis. For four QTLs, we identified missense mutations in genes known to be involved in meiotic recombination among the most significantly associated variants. Most of the identified mutations had significant effects in both genders with three of them accounting each for approximately 10% of the genetic variance in males (the allelic substitution effect being approximately equal to one additional CO per genome). Thus, a large fraction of the genetic variance is associated with missense mutations in genes known to be involved in meiotic recombination. Our results are very different from reports of recombination in other species. For instance, in human, recombination rate is higher in females, distinct variants affect recombination rate in males and females, and the genetic correlation is close to 0, whereas in cattle, we observed a higher recombination rate in males controlled by shared variants effective in both sexes.

5

Genetic covariance between additive and dominant effects in closed lines selected for more than 40 generations

Eduardo N Fernandez[1], Andres Legarra[2], Juan P Sanchez[3], Ruben D Martinez[1], Manuel Baselga[4]
[1]Universidad Nacional de Lomas de Zamora, [2]INRA, [3]IRTA, [4]Universidad Politecnica de Valencia

Inbreeding leads to deviations from Hardy-Weinberg equilibrium, which create a correlation between breeding values and dominant deviations. In addition, dominance variation can be split into “inbred” and “outbred” parts. The value of these variances and covariance sheds light on the genetic architecture of the traits and also on the outcome of evolutionary processes in small, inbred populations. Estimation of these parameters has rarely been done satisfactorily due to its complexity.

We estimate variance and covariance components for additive and dominance effects affecting the selected trait “litter size at weaning” in three closed populations of rabbits raised at the Universitat Politecnica de Valencia, Spain: lines A, V and H with 6443, 6689 and 2391 animals spanning 43, 39 and 15 generations of selection respectively. Inbreeding cumulates at a rate of ~0.8% per generation. Pedigrees and phenotypes are complete.

Based on pedigree, we computed the probabilities of the nine condensed identity coefficients using Karigl’s (1981), then we used Bayesian methods by Gibbs Sampler to obtain estimates, using mixed model equations equivalent to DeBoer and Hoeschele (1993). Model included inbreeding depression. We give expressions for estimators of variance and covariance components in the last generations as opposed to the base populations.

In the last generations, the estimated correlation between additive effects and dominant deviations is -0.3 for lines A and V and 0 for line H. Inbred dominant variation becomes increasingly important (in line H is even higher that additive variation) so that total genetic variation increases in line H and is almost constant for lines A and V in spite of a decrease of ~40% of additive variance since the foundation of these lines.

6

Accounting for randomness of genotypes in across and single population genome-enabled prediction: A hierarchical Bayes approach

Mr. Carlos A. Martinez[1,2], Dr. Kshitij Khare[2], Dr. Arunava Banerjee[3], Dr. Mauricio A. Elzo[1]
[1]Department of Animal Sciences, University of Florida, [2]Department of Statistics, University of Florida, [3]Department of Computer and Information Science and Engineering, University of Florida

Multiple linear regression models used in genome-enabled prediction ignore randomness of genotypes. Therefore, analyses are conditional on observed genotypes. Furthermore, most across population genome-enabled prediction studies are carried out by pooling data and fitting models developed for single population analyses. In this study, a family of models to perform single and across population genome-wide prediction modeling genotypes as random variables and allowing population-specific effects for each marker (for across population analysis) was developed using a hierarchical Bayes approach. Models differed in the priors used and assumed residual variances to be either homogeneous or heterogeneous. To account for randomness of genotypes, the joint probability mass function of marker genotypes conditional on allelic frequencies and pedigree information was derived. Thus, these models incorporated kinship and genotypic information that not only permitted to account for heterogeneity of allelic frequencies in across population studies, but also to include individuals with missing genotypes at some or all loci without the need for previous imputation. This was achieved by treating the non-observed genotypes as unknown model parameters. For the across population case, Bayes factors and fractional Bayes factors to compare models with their null versions (models ignoring population structure, but still accounting for randomness of genotypes) were derived via the Laplace approximation. Implementation of these models and computation of model comparison criteria were illustrated using simulated data. Theoretical and computational issues as well as possible extensions and refinements were discussed. Features of this set of models make them promising for genome-enabled prediction. Inclusion of information from the probability distribution of genotypes is perhaps the most attractive. Further research assessing the performance of this family of models and comparing them with conventional models used in genome-enabled prediction is needed.

7

Dissecting the genetic and environmental contributions to anthropometric and cardio-metabolic trait variation in humans.

Mr. Charley Xia[1], Dr Carmen Amador[1], Dr. Archie Campbell[2], Prof. David Porteous[2], Prof. Nicholas D. Hastie[1], Dr. Caroline Hayward[1], Dr. Veronique Vitart[1], Dr. Pau Navarro[1], Prof. Chris S. Haley[1,3]
[1]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, [2]Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, [3]The Roslin Institute and R(D)SVS, University of Edinburgh

Genome-wide association studies have successfully identified thousands of loci affecting a range of human complex traits and diseases. The proportion of phenotypic variance explained by significant associations is, however, limited. Given the same dense SNP panels, mixed model analyses capture a greater proportion of phenotypic variance than single SNP analyses but the total is generally still less than the genetic variance estimated from pedigree studies. Combining information from pedigree relationships and SNPs, we analysed 16 anthropometric traits in a Scottish population comprising ~20,000 individuals -that were part of Generation Scotland's Scottish Family Health Study- genotyped for ~520,000 common autosomal SNPs. We found that, on average across traits, SNP-associated and pedigree-associated genetic effects each explained around half the genetic variance with recently-shared environment of couples accounting for ~11% of the phenotypic variance. The joint family environment shared by parents and offspring had a limited impact on trait variation as did the past environment shared by siblings. Our findings point to appropriate models to use in future studies as pedigree-associated genetics and couple environment have seldom been taken into account in genotype-based analyses. Appropriate description of the trait variation could help understand causes of inter-individual variation and facilitate the detection of contributing loci and environmental factors. Even with a large number of common SNPs, much of the genetic variance is not captured by these but by pedigree-associated effects (that may reflect rare genetic variants), providing insight on potential trait architectures. In samples of unrelated individuals commonly used in GWAS, the pedigree-associated effects we captured here cannot be captured, and could explain at least part of the missing heritability.

8

Effect of genetic architecture and sample size on the accuracy of genomic prediction of complex traits

Fabio Morgante[1], Dr Wen Huang[1], Dr Christian Maltecca[2], Dr Trudy FC Mackay[1]
[1]Program in Genetics, Department of Biological Sciences, and WM Keck Center for Behavioral Biology, North Carolina State University, [2]Program in Genetics, and Department of Animal Science, North Carolina State University

Understanding the genetic architecture of complex traits is a fundamental aim of many branches of genetics. Genome wide association studies (GWAS) have been successful at identifying some loci affecting complex traits. However, those loci account for just a very small proportion of the total genetic variation, the widely known phenomenon called “missing heritability”. As a result, prediction of phenotypes based on the loci uncovered by GWAS has had low accuracy. Methods that regress phenotypes on hundreds of thousands of markers concurrently (whole genome regression, WGR,) may be able to capture a conspicuous amount of the genetic variation of complex traits and, thereby, increase predictive ability. However, most WGR methods assume strict additivity and traits that are potentially affected by non-additive (epistatic) interactions have not shown any gain in predictive ability. Here, we investigated the effect of sample size and genetic architecture on the accuracy of genomic prediction of complex traits. We used G-BLUP methodology and the unique resource of the Drosophila Genetic Reference Panel (DGRP), a collection of 205 fully sequenced inbred lines that have been phenotyped for many quantitative traits, as well as simulated data. The results show that the accuracy of prediction increases as the sample size increases, conditional on the genetic architecture of the trait examined being taken into account in the statistical model used. In particular, strict additive models in presence of an epistatic component being part of the genetic architecture may fail completely, even if we have knowledge about the true variants affecting that trait and no matter the size of the sample. However, when an epistatic model is fitted, the accuracy of prediction rises, even with small sample size. In summary, this study shows the importance of accounting for their genetic architecture to increase the accuracy of genomic prediction of complex traits.

9

Functional validation of loci contributing to nicotine resistance in Drosophila

Mr. Chad A. Highfill[1], Miss. Samantha K.T. Nguyen[1], Mr. Jonny H. Tran[1], Mrs. Xiaofei Wang[1], Miss. Taylor R. Moldenhauer[1], Dr. Stuart J. Macdonald[1]
[1]*University of Kansas*

All species posses detoxification pathways enabling plastic responds to environmental toxins. Given the widespread use of pesticides and frequent development of resistance to these compounds by crop pests, it is critical to understand the genetic basis of xenobiotic resistance in insects. Previously, we used the Drosophila Synthetic Population Resource (DSPR) to map four QTL contributing to resistance to nicotine, a compound employed as a natural insecticide by some plants. RNA-seq showed that both cytochrome P450 genes under QTL1, Cyp28d1 and Cyp28d2, and one of ten UDP-glucuronosyltransferase (Ugt) genes under QTL4, Ugt86Dd, were differentially-expressed between susceptible and resistant strains. Here, we carry out RNAi, quantitative complementation tests (QCTs), overexpression experiments, and mutant analysis to functionally validate the effects of these loci on nicotine resistance, and attempt to identify causative sequence variants. Ubiquitous and targeted RNAi depletion of Cyp28d1 and Ugt86Dd reduced nicotine resistance, providing evidence these genes are involved in the phenotype. To determine whether these loci harbor segregating variation influencing resistance, we crossed susceptible and resistant strains to a range of deficiencies and insertional mutants. These QCTs revealed functional allelic variation at Cyp28d1 and Cyp28d2, and suggested that multiple nicotine resistance factors are present under QTL4, consistent with the idea that several of the Ugt genes are involved. Sequencing the Ugt86Dd open reading frame revealed a 22bp coding deletion segregating in the DSPR, with the four most susceptible founder haplotypes at the QTL all harboring the deletion allele. We constructed overexpression genotypes using both Ugt86Dd alleles, and found that nicotine resistance is significantly greater following anterior midgut overexpression of the insertion allele compared to the deletion allele, implying the variant has a functional role. We have now successfully generated custom Ugt86Dd mutants using the CRISPR-Cas9 system, and future experiments will help elucidate the effect of the locus on nicotine resistance.

10

GGE biplot of genotype by environment interaction for tuber yield of potato clones resistant to potato virus Y

Mr. Mario Henrique Murad Leite Andrade[1], Dr Cesar Augusto Brasil Pereira Pinto[1]
[1]*UFLA - Universidade Federal de Lavras*

The genotype by environment interaction is a reality for breeders and farmers and one of the biggest obstacles for breeding programs, since its disregard can over or under estimate the genotypic value. This topic has been extensively studied, especially in tropical conditions, where there are more heterogeneous environments, which may result in higher interaction. Potato (Solanum tuberosum L.) is an important crop in Minas Gerais State that accommodate 30% of the planted area in Brazil. There are not varieties widely adapted to tropical conditions and all cultivars are susceptible to Potato Virus Y (PVY), causing up to 100% loss in tubers production. The aim of this study was to evaluate the adaptability and stability of potato clones resistant to PVY in various environmental conditions in southern Minas Gerais using the GGE Biplot method. Eighteen clones, from the potato breeding program of UFLA, possessing the allele Ryadg, which confers resistance to PVY, with high total production and good tuber appearance were pre-selected and used in the experiment together with four control cultivars usually planted in the state. They were evaluated at six locations in 2014/2015, in potato growing regions of Minas Gerais State, using RCB design. Combined ANOVA indicated that the main effects of genotypes, locations and GxE were significant. GxE was the largest source of variation representing 36% of total variation. The two first principal components were used for the construction of Biplot and explained 44.8% and 25.8% respectively of the total sum of square of the interaction. Biplot showed that the clones are more adapted and stable compared with the cultivars and one of them (MLG-23.37) stands out for its second best average combined with high stability. Its cultivation in the southern State of Minas Gerais is indicated.
Key words: potato breeding; genotype x environment interaction; GGE Biplot, PVY.

11

The use of marker-data transformations to account for linkage disequilibrium in genomic selection: a case study in switchgrass (Panicum virgatum L.)

Mr. Guillaume P Ramstein[1], Dr. Joseph Evans[2], Dr. Shawn M Kaeppler[3], Dr. Robert B Mitchell[4], Dr. Kenneth P Vogel[4], Dr. C Robin Buell[2], Dr. Michael D Casler[5]

[1]Department of Agronomy, University of Wisconsin-Madison, [2]Department of Plant Biology and U.S. Department of Energy Great Lakes Bioenergy Research Center, Michigan State University, [3]Department of Agronomy and U.S. Department of Energy Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, [4]Grain, Forage, and Bioenergy Research Unit and Agricultural Research Service, United States Department of Agriculture, University of Nebraska, [5]Department of Agronomy, University of Wisconsin-Madison, and Agricultural Research Service, United States Department of Agriculture

Genomic selection (GS) is an attractive technology to generate rapid genetic gains, particularly in perennial grass species like switchgrass, where phenotyping generally requires at least two years of field trial. In this study, we empirically assessed prediction procedures for GS 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 141,030 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 account of linkage disequilibrium among markers, offer valuable opportunities for improving prediction procedures in GS. Furthermore, our analyses indicate that the cases in which marker-data transformations seem advantageous are those where the causal loci are underrepresented in the marker data, thereby bringing insight into why and when accounting for linkage disequilibrium might be useful in genomic prediction.

12

Quantitative exploration of Ethiopian durum wheat: joining genetics and social sciences to produce a better wheat through GWAS and a multiparental population design

Dr. Matteo Dell'Acqua[1], Dr. Yosef G. Kidane[1,2,3], Dr. Dejene K Mengistu[1,4], Miss. Chiara Mancini[1], Dr. Elisabetta Frascaroli[5], Dr. Carlo Fadda[2], Prof. Mario Enrico Pe'[1]
[1]Scuola Superiore Sant'Anna, [2]Bioversity International, [3]Sirinka Agricultural Research Center, [4]Makelle University, [5]University of Bologna

Ethiopia is one of the biggest and most populous countries in Africa, counting 100 millions of inhabitants, three-quarters of whom employed in subsistence agriculture. The Ethiopian plateau is a center of diversity for tetraploid wheat, a staple food for smallholder farmers and a high-value crop used to for bread, pasta, and semolina preparations worldwide. We explore the quantitative genetics of Ethiopian wheat through a multi-faceted approach spanning agronomy, genetics and farmers' knowledge elicitation. We collected 400 wheat landraces representing Ethiopian durum wheat diversity and characterized their molecular diversity for more than 80K SNP loci, the state of the art of wheat genotyping. We found that the genetic makeup of Ethiopian wheat is completely different from the international allele pool, confirming its uniqueness. We selected 50 Ethiopian durum wheat landraces maximizing diversity and we intercrossed them with a recurrent improved line, Asassa, to produce the first nested association mapping (NAM) population in durum wheat, now comprising more than 7,000 recombinant inbred lines (RIL) at F6. We genotyped 1,200 NAM RIL and we report on their structure and diversity. Concurrently, we tested the full collection of 400 landraces in two environments in the Ethiopian highlands, collecting data for 10 traits of agronomic interest. We conducted a genome wide association (GWA) study identifying several loci with potential in international wheat breeding. Two wheat growing smallholder farming communities of 30 persons each were organized in gender-uniform groups and asked to evaluate the 400 wheat genotypes for four key quality traits identified through focus group discussions: earliness, tillering capacity, spike quality and overall appreciation. We coupled this information with metric measures of phenotypes breaking down farmers' appreciation of wheat varieties, identifying the best trait combinations. We propose to use this information to conduct GWA on farmers' appreciation and identify breeding targets addressing smallholder farmer needs.

13

Asymptotic normality of breeding values under segmental inheritance.

Dr Rodolfo Cantet[1], Dr Sebastián Munilla[1]
[1]*University of Buenos Aires,* [2]*INPA Conicet*

Normality in the infinitesimal additive model is based on the assumption of a large number of loci segregating independently across the genome. Thus, depending on parental inbreeding Mendelian residuals explain somewhat less than half the additive variance. To increase the accuracy of predictions of breeding value, genomic markers bring extra-information that reduce Mendelian residual variance and allow detecting realized relationships rather than those expected by the pedigree. When looking at the inheritance process in diploids in which segments of genomes are passed rather single loci, genomic markers are informative sources of recombinations with respect to the expected proportions of ancestral genomes received by an individual from ancestral generations. We model the inheritance of segments using R. A. Fisher's theory of junctions and P. Stam's number of segments (Ns) in the population. The variance of a segment involves two parameters: the additive variance and the covariance between additive effects from loci in linkage disequilibrium. Whereas covariances of segments involves the same two parameters, specification of the joint distribution of segments within and across individuals requires the two-locus theory of C. C. Cockerham and B. S. Weir involving digametic, trigametic, and tetragametic probabilities. All of these probabilities are functions of the product of gene frequencies under linkage equilibrium, plus terms that are functions of linkage disequilibrium measures among two, three or four alleles. These latter terms are also functions of (1/Ns). Hence, as Ns gets large the terms involving linkage disequilibrium measures go to zero and so does the covariance between additive effects under linkage disequilibrium. This argument is used to prove the asymptotic normality of the breeding values under segmental inheritance of the additive two-locus model and, in separate research, the covariance of breeding values using pedigree plus genomic markers.

14

Inbreeding coefficient under a grandparental regression model

Dr. Sebastián Munilla[1], Dr. Rodolfo J. C. Cantet[1,2]
[1]*University of Buenos Aires,* [2]*INPA - CONICET*

The linear regression of the breeding value of an individual on his parents' breeding values is termed the 'parental regression'. Under this representation, the regression coefficients are equal to one half and the regression error term is called the Mendelian residual. If the parents are not inbred, the variance of the Mendelian residuals equals to half the additive variance. More generally, the variance is a function of the parent's inbreeding coefficients and the additive variance. Mendelian residuals are interpreted as terms that reflect uncertainty due to not knowing which fractions of grandparental genome were passed to the individual over or below one-fourth. Genomic information, however, may help disentangling the realized inheritance process. To take advantage of this fact, we expanded the parental regression model with a linear combination of breeding values from all four grandparents, and termed this new representation as 'grandparental regression model'. The new regression coefficients are related to the fraction of genome shared identical by descent between the individual and the corresponding grandparent. Precisely, they are defined as half the difference between the genome shared identical by descent between the individual and their paternal (or maternal) grandparents due to recombination. The halving is due to segregation of the grandparental chromosomes into the parental gamete that originated the zygote. By operating algebraically on this model we provide formulae for the inbreeding coefficient and the variance of the Mendelian residuals. The resulting inbreeding coefficient is a composition, over each common ancestor that connects both parents, of three terms: 1. the parent's expected relationship as defined by the genealogy; 2. a product of the grand parental regression coefficients; 3. a combination of both, resulting from counting all possible paths connecting the parents to the common ancestor under a directed acyclic graph representation.

15

The genetic basis of evolved acute stress response in the nematode Caenorhabditis remanei

Ms. Christine H O'Connor[1], Dr. Kristin L. Sikkink[2], Dr. Janna L Fierst[3], Dr. John Willis[1], Dr Patrick C Phillips[1]
[1]*University of Oregon,* [2]*University of Minnesota ,* [3]*The University of Alabama*

Stress response is a functionally important trait whose genetic basis is thought to be conserved throughout animals. Single gene knockout studies and gene expression analysis of the classic model nematode Caenorhabditis elegans have identified gene pathways though to be important the response to heat, oxidative and other abiotic stresses, but the genetic basis of heat stress and oxidative stress response has yet to be examined within an experimental evolution framework. We are investigating the genetic basis of acute heat stress and acute oxidative stress response in a set of experimentally evolved populations of the polymorphic nematode C. remanei. Both heat and oxidative stress are thought to cause an accumulation of misfolded proteins and activation of the insulin-like signaling stress response pathway. However, phenotypic data from experimentally evolved populations shows that adaptation to heat stress does not have a pleiotropic effect on oxidative stress resistance and vice-versa. We use whole genome sequencing and classic population genetics statistics to locate signatures of selection in the genome, comparing the ancestral population with the evolved and lab adapted populations to test whether the lack of phenotypic pleiotropy to stress response is reflected at a the genotypic level. This extensive set of genomic data, coupled with a fully annotated reference genome, helps to specify the size, location and number of regions under selection, allowing us to identify genomic patterns of selection for a multifactorial trait, and information about the genetic basis of heat stress and oxidative stress response. In addition, analysis of transcriptome data from the same set of populations allows us to correlate regions of putative selection with gene expression data, and identify genes whose expression changes are affected by cis or trans-factor mutations.

16

Genome-wide association studies of additive and dominance effects for female fertility traits in Danish Holstein cattle

Mr. Xiaowei Mao[1,2], Dr. Goutam Sahana[1], Dr. Anna M Johansson[2], Dr. Dirk-Jan De Koning[2], Dr. Bernt Guldbrandtsen[1]
[1]*Aarhus University,* [2]*Swedish University of Agricultural Sciences*

Intensive selection on yield in dairy cows has led to a decline in fertility, because of unfavorable genetic correlations between yield and fertility traits. Declined fertility has increased costs due to extra inseminations, veterinary treatments, and replacement cows. Most gene mapping studies in cattle do not consider dominance effects in the analyses, mainly because the analyses were based on genotyped bulls while the phenotypes were derived from their daughters' performance. With large datasets routinely collected as part of the modern precision dairy farming, it is possible to identify regions in the taurine genome that exhibit signs of dominance. The objectives of this study were to detect loci with additive and/or dominance effects on female fertility traits in Danish Holstein cattle. Analyses were carried out on several fertility traits: number of inseminations per conception (AIS), days from calving to first insemination (ICF) (only for cows), days from the first to last insemination (IFL), and gestation length (ILC) for both heifers (2,664) and cows (2,972) genotyped using Illumina BovineSNP50 array. There were 38,321 SNPs left after the quality control. A genome-wide association study was performed by fitting the single SNP (Single Nucleotide Polymorphism) in a linear mixed model with both an additive and a dominant effect. SNPs with a significant genetic association were tested for their mode of gene action (additive or dominant). For heifers, we identified one quantitative trait locus (QTL) on Bos taurus autosome (BTA) 5 for IFL and ILC. For cows, we identified one QTL on BTA 3 for ICF and one QTL on BTA 9 for ILC. In general, additive effects explained more genetic variance than dominance effects. Our results showed that dominance effects play a relevant role in the genetic architecture of female fertility traits in cattle.

17

Bagging elastic net

Alencar Xavier[1], Bill Muir[1], Katy Rainey[1]
[1]*Purdue University*

Genome-wide selection represents the next-step for genetic improvement of plants and animals. Whole-genome regression (WGR) models in Bayesian framework represent the main family of prediction methods, and there exist much space for improvement. Over the past decade, researchers have tried to overcome modeling pitfalls for hyper-dimensional data with flexible assumptions and efficient algorithms.

Major changes observed overtime in prediction models regard types of kernels and the distribution of marker effects, dealing shrinkage and variable selection. However, the way we define the model's loss function has not change. WGR methods use L2 norm, with the exception of LASSO (L1 norm). Elastic net (EN) was proposed to overcome the limitation of both. This type of model can produce L1 and L2 coefficients, controlled by a parameter alpha.

Re-sampling techniques can be used to increase accuracy while decreasing bias and computing time. In the proposed ensemble method, we split the data into two fractions, A and B, in each MCMC interaction. Regression coefficients are updated using the fraction A with EN method, whereas fraction B is used to determine out-of-bag prediction parameters. With bagging, the parameter alpha can then found using the Metropolis algorithm, using fitness and OOB parameters to accept values.

Using datasets from SoyNAM and CYMMIT, two soybean traits and four wheat yield environments, we evaluated the performance of a Bagging EN model, tracking the alpha parameter via Metropolis, in comparison to LASSO (L1 loss) and Ridge regression (L2 loss). We concluded that EN loss function is beneficial to genomic prediction.

18

When one plus one does not equal two: some tandem gene duplicates are overactive

David W. Loehlin[1], Sean B. Carroll[1]
[1]*UW-Madison and HHMI*

Tandem gene duplication is an important mutational process in evolutionary adaptation and human disease. Hypothetically, two tandem gene copies should produce twice the output of a single gene, but this expectation has not been rigorously investigated. Here, we show that tandem duplication often results in more than double the gene activity. A naturally occurring tandem duplication of the Alcohol dehydrogenase (Adh) gene exhibits 2.6-fold greater expression than the single copy gene in transgenic Drosophila. This tandem duplication also exhibits greater activity than two copies of the gene in trans, demonstrating that it is the tandem arrangement and not copy number that is the cause of overactivity. We also show that tandem duplication of an unrelated synthetic reporter gene is overactive (2.3- to 5.1-fold) at all sites in the genome that we tested, suggesting that overactivity could be a general property of tandem gene duplicates. Overactivity occurs at the level of RNA transcription, and therefore tandem duplicate overactivity appears to be a novel form of position effect. The increment of surplus gene expression observed is comparable to many regulatory mutations fixed in nature, and if typical of other genomes, would shape the fate of tandem duplicates in evolution.

19

Prediction of top cross performance in maize for drought tolerance via genome-wide markers

Dr Luciano D. C. E Silva[1], Mr Kaio O. D. G. Dias[2], Dr Maria Marta Pastina[3], Dr Claudia T. Guimarães[3], Dr Sidney N. Parentoni[3], Dr Paulo E. D. O. Guimarães[3], Dr Arley F. Portugal[3], Dr Newton P. Carneiro[3], Dr Roberto Noda[3], Dr Kelci J. Miclaus[1], Dr Russ D. Wolfinger[1], Dr Lauro J. M. Guimarães[3]
[1]SAS Institute Inc., [2]Federal University of Lavras, [3]Embrapa Maize and Sorghum

In a maize breeding program, there is a continuous production of advanced inbred lines; but, because the commercial maize market is based in hybrids, business interests favor evaluation of the performance of hybrids, rather than the lines on their own merit (per se). However, making all possible crosses among hundreds (or thousands) of lines and evaluating their hybrids in field experiments is extremely costly and time consuming, if not infeasible. Instead, some pre-selected set of lines (testers) known to produce high performance hybrids are crossed with new lines in an array known as a top cross. The maize breeding program of Embrapa Maize and Sorghum in Brazil has a set of 179 lines and 2 testers that were field-evaluated on a per se basis under water-stressed (WS) and well-watered (WW) conditions. A subset of 160 of these lines were crossed with the testers, and their hybrids were evaluated under WS and WW conditions. Lines were genotyped via GBS, producing 57047 SNPs, after quality control processing. Here, we rank-compare in silico top crosses as predicted by a linear model fitted on the per se yield phenotype and SNP markers data, against the field-evaluated hybrids from the corresponding top crosses. The number of matching hybrids among the top 30 field-evaluated and top 30 in silico were 8 and 7 for WS and WW, respectively. Additionally, we made all possible in silico crosses (16.290) among the 179 lines, and showed that some drought-tolerant lines produced high performance hybrids on a broad range of crosses, which was previously unknown. The results also show that the in silico prediction has great potential in identifying new crosses for the extraction of new superior lines within heterotic groups. Such an in silico prediction approach has huge potential to improve the speed and effectiveness of breeding programs.

20

Holistic selection for eggshell quality using a bio-complex in layer chicken

Dr. Jesus Arango[1], Dr. Petek Settar[1], Dr. Anna Wolc[1,2], Dr. Neil P. O'Sullivan[1]
[1]*Hy-line International,* [2]*Department of Animal Sciences, Iowa State University*

Shell quality is of paramount importance in layers. Traditionally, selection for shell quality was based on static measurements using destructive methods, such as breaking strength (BS) or non destructive ones. A recent dynamic method, the acoustic egg test (AET) combines vibration profile of the shell with egg mass to calculate the dynamic stiffness (KDyn), a quantitative indicator of integral shell resistance. Acoustic profiles allow threshold detection of micro-cracks (MCR). A shell quality bio-complex was defined to improve overall shell quality. Traits measured were: BS at equator (BSE) and poles (BSP), KDyn and MCR (as continuous or categorical), taken at 4 different ages (26, 42, 65 and 84 wk) on multiple eggs/hen-age. A total 89845 records from a RIR line were evaluated. Two models were implemented: I) four-trait linear repeatability model with fixed effects of contemporary group, lab station and the covariates of age and egg weight; and random additive genetic and permanent environment effects and II) three-trait linear (BS, KDyn)-threshold (MCR) model, including type of measurement (E, P). Model (I) was implemented with AIREMLF90 and (II) with THRGIBBS1F90, with a 100,000 Gibbs chain (5,000 as burn-in). Heritability and repeatability estimates (I) were: h^2= 0.14, 0.20, 0.34 and 0.02; r= 0.15, 0.34, 0.44 and 0.03 for BSE, BSP, KDyn and MCR, respectively. Genetic correlation between BSE and BSP was 0.60, the BS traits were moderately correlated with KDyn (0.29, 0.41), and negatively correlated with MCR (-0.28, -0.56). KDyn and MCR were negatively correlated (-0.50). Model II had similar results, except for increased h^2 (0.06) and r(0.09) for MCR. Results indicate that BSE and BSP are different traits; while incidence of MCR is lowly heritable but significantly negatively correlated with the other three traits, favoring an indirect selection of MCR via BSE, BSP and KDyn for a holistic selection to improve shell quality.

21

Genome-wide and species-wide variation in C. elegans reveals association of telomere length with population differences in pot-2

Daniel E. Cook[1,2], Stefan Zdraljevic[1,2], Robyn E. Tanny[1], Erik C. Andersen[1,2]
[1]Dept. of Molecular Biosciences, Northwestern University, [2]Interdisciplinary Biological Sciences Program, Northwestern University

Model organisms have greatly advanced our understanding of biological phenomena. However, experiments in model organisms are frequently carried out in a single laboratory strain (e.g. N2 in C. elegans). This approach, although useful for many applications, neglects insights that can be made by examination of genetic diversity in natural populations. This largely untapped resource can help tease apart the genetic factors responsible for complex traits and address questions concerning etiology of disease, genome evolution, and population genetics.

Given the benefits of studying natural populations, we have sequenced the whole genomes of 211 wild isolate strains of C. elegans that represent 152 distinct genome-wide haplotypes. To date, this collection is the largest set of wild isolate sequences in C. elegans. Sequences were aligned and achieved a median coverage of 72x, giving us the ability to identify rare and common variants with high confidence. Across all of these strains, we called approximately 1.3 million single nucleotide variants that are predictive of functional changes in C. elegans.

Using this dataset, we mapped trait differences ascertainable from sequence data, including estimates of telomere length. We found that telomere length is strongly associated with a region on chromosome II that contains the gene pot-2 (Protection of Telomeres). POT-2 represses the telomere-lengthening enzyme telomerase. We identified a nonsynonymous (F68I) variant within the OB-fold domain of POT-2 that is correlated with longer telomeres. Additional analysis on mutagenized strains revealed enrichment for mutations specifically within the OB-fold of POT-2 in long-telomere strains. We also examined whether telomere length affected brood size, longevity, or the germlines of strains and did not observe any correlations. Our results suggest that telomere length does not correlate with fitness. Additionally, the F68I variant is not under selection in wild populations, indicating that variation in telomere length among individuals likely has little appreciable benefit.

22

Natural variation alters sensitivities to topoisomerase II poisons through changes in conserved drug-binding residues

Stefan Zdraljevic[1], Robyn E. Tanny[1], Tyler C. Shimko[1], Erik C. Andersen[1]
[1]*Molecular Biosciences, Northwestern University, Evanston, IL, United States*

Individuals in a population vary in their responses to chemotherapeutic interventions. The identification of the genetic determinants underlying this variability has the potential to improve the administration and efficacy of current drugs while directing the development of new drugs tailored to specific variants. However, it remains difficult to identify genetic variants that are predictive of therapeutic efficacy in human studies because of high genotyping costs, inability to control environmental conditions, and the polygenic nature of a majority of biomedically relevant traits. To circumvent these issues, we used the powerful model Caenorhabditis elegans to identify 40 quantitative trait loci (QTL) that underlie differences in susceptibility to six topoisomerase II poisons commonly used in chemotherapeutic regimens to treat various cancers. A scan of genetic variants underlying one of these QTL identified the candidate gene top-2, which encodes for one of the two known topoisomerase II proteins in the C. elegans genome. Strains sensitive to the poisons contain a top-2 allele with several predicted synonymous and non-synonymous variants when compared to the standard laboratory strain N2. A top-2 deletion in the N2 genetic background failed to complement the sensitive top-2 variant phenotype, suggesting that top-2 is the causal gene underlying this QTL. We hypothesize that the Q797M variant present in sensitive strains leads to more stable binding of etoposide to TOPOII and therefore increases its potency as a poison. Interestingly, the two human versions of TOP2 (alpha and beta) vary at this residue and have differential binding affinities to one of the poisons we tested. Our identification of conserved drug susceptibilities between humans and C. elegans along with our ability to probe the genetic determinants of these susceptibilities has introduced C. elegans as a powerful model for elucidating the mechanisms underlying population variation in chemotherapeutic drug responses.

23

Genetic control of fumonisin incidences in tropical maize

Dr. Joao Souza[1], Dr. José Maria Villela Pádua, Dra. Maria Marta Pastina, Ms Kaio Olimpio Graças Dias, Dr Lauro Jose Moreira, Mrs. Maria Beatriz Silva
[1]*Federal University of Lavras*

The occurrence of Fusarium ear rot in maize cause not only direct losses with yield and grain quality reductions but also indirect losses due to the fact these substances are toxic to humans and animals. Despite the importance of fumonisins, there are no knowledge about the genetic control of these mycotoxins in tropical maize. Thus, this study had the aim to investigate the genetic control of the incidence of fumonisins in tropical maize through a complete diallel, via linear mixed model. For this purpose, 13 contrasting inbreed lines were identified, considering different heterotic groups. Single-cross hybrids were assessed in three experiments, using a randomized complete block design with two replications. The incidence of fumonisins in maize kernels was evaluated in laboratory according to the protocol proposed by Vicam (1998). Based on the mixed model approach, individual, for each experiment, and joint diallel analyses were performed. In the joint diallel analysis, different models for the genetic (G) and residual (R) variance-covariance matrices were examined. Moreover, the additive genomic relationship information among inbreed lines were also included in those models. Experimental accuracies, measured by the index of selective accuracy, were considered adequate in all analyses. Based on the results of this study, it was possible to infer that there is a predominance of non-additive effects in the genetic control of the incidence of fumonisins in tropical maize.

25

Does selection act on gene networks? A case study in the model plant Medicago truncatula.

Mr Maxime Bonhomme[1], Ms Mathilde Negretto[1], Mr Bernard Dumas[1], Mr Christophe Jacquet[1]
[1]*Université de Toulouse, Paul Sabatier*

Medicago truncatula is a model legume species used to investigate the genomics of plant–microbe interactions, notably of root symbioses. In a recent study (Bonhomme et al. 2015, Mol Biol Evol), we have performed a fine-scale genome scan of selective sweep signatures in M. truncatula using more than 15 million single nucleotide polymorphisms – SNP - identified on 283 natural inbred accessions in the frame of the Medicago Hapmap Project (http://www.medicagohapmap.org/). By mining functional annotation and available expression data of candidate genes, we concluded that root-associated microorganisms were the main selective agents driving long-term adaptation in Medicago truncatula. Yet, a striking result was that some candidate genes showed strong correlation in their patterns of gene expression, suggesting that co-expressed gene clusters might be the target of epistatic selection.

To investigate the conditions required for epistatic selection to be detected with SNP data, we developed a simulation pipeline to study the evolution of allele frequencies at two physically unlinked loci under neutrality or positive epistatic selection. Preliminary results indicate that across a range of conditions the fixation rate of coadapted allelic combinations is higher in inbred species compared to outcrossers. In addition, we observed that linkage disequilibrium between two physically unlinked genes under epistatic selection can be strong and thus be taken as evidence for gene coadaptation. To illustrate this, we calculated pairwise mean LD values in a set of genes containing candidate genes identified in Bonhomme et al. (2015) plus genes belonging to the Common Symbiosis Signalling Pathway (CSSP) in M. truncatula. The most extreme LD values were observed between three genes: a Serine/Threonine kinase, a Cysteine-rich receptor-kinase-like and a LysM Receptor-Like Kinase of the CSSP ($0.93 < D' < 0.97$). These results highlight the potential to detect networks of coadapted genes by searching for epistatic selection signals using SNP data.

26

Covariance association test (CVAT) identify genetic markers associated with schizophrenia in functionally associated biological processes

Palle Duun Rohde[1,2,3], Ditte Demontis[2,3,4], Beatriz Castro Dias Cuyabano[1], The GEMS Group[5], Anders D. Børglum[2,3,4], Peter Sørensen[1]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, [2]Centre for Integrative Sequencing, iSEQ, Aarhus University, [3]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, [4]Department of Biomedicine, Aarhus University, [5]The GEMS (Genomic Medicine for Schizophrenia) Group

Schizophrenia is a psychiatric disorder with large personal and social costs, and understanding the genetic etiology of this disorder is important. Such knowledge can be obtained by testing the association between a disease phenotype and individual genetic markers. However, such single marker methods have limited power to detect genetic markers with small effect sizes. Aggregating genetic markers based on biological information might increase the power to identify sets of genetic markers of etiological significance. Several set test methods, and here we propose a new set test method derived from genomic best linear unbiased prediction (GBLUP): the Covariance Association Test (CVAT). CVAT measures the covariance between the joint genetic values for the genetic markers in the feature set and the genetic values for the remaining genetic markers, outside the feature set. A feature set is a group of genetic markers defined by biological information, e.g. genes and pathways. Because CVAT was derived from GBLUP, many extensions exist. One such is to allow the genetic marker effects to be a mixture distribution. Such mixture distribution can be obtained by adding an extra genetic variance component to the GBLUP. The two genetic variance components contained either the genetic effects of rare or common genetic markers. We compared the performance of CVAT to other commonly used set test methods using a simulated population with the same genetic parameters as schizophrenia. We found that CVAT was among the top performer. When extending CVAT to utilize a mixture of SNP effects, we found an increase in power to detect the causal sets of genetic markers. Applying the methods to a Danish schizophrenia case-control data set, we found genomic evidence for association of schizophrenia with vitamin A metabolism and immunological responses, which previously have been implicated with schizophrenia based on experimental and observational studies.

27

Systematic bias of correlation coefficient may explain negative accuracy of genomic prediction

Mr. Yao Zhou[1,2], Dr. Zhiwu Zhang[1,2]
[1]*Department of Crop and Soil Sciences, Washington State University,* [2]*College of Life Science, Northeast Agricultural University*

Accuracy of genomic prediction is commonly calculated as the Pearson correlation coefficient between the predicted and observed phenotypes in the inference population by using cross-validation analysis. More frequently than expected, significant negative accuracies of genomic prediction have been reported in genomic selection studies. These negative values are surprising, given that the minimum prediction of accuracy should hover around zero when randomly permuted datasets are analyzed. We reviewed the two common approaches for the calculation of Pearson correlation and made the hypothesis that these negative accuracy values reflect potential bias due to artifacts caused by the mathematical formulas used to calculate prediction accuracy. The first approach, Instant accuracy, calculates correlations for each fold and reports prediction accuracy as the mean of correlations across fold. The other one, Hold accuracy, predicts all phenotypes in all fold and calculates correlation between the observed and predicted phenotypes at the end of the cross-validation process. Using simulated and real data, we demonstrated that our hypothesis is true. Both approaches are biased downward under certain conditions. The biases become larger when more fold are employed and when the expected accuracy is low. The bias of Instant accuracy can be corrected using a modified formula.

28

Bayesian-information and Linkage-disequilibrium iteratively nested keyway (BLINK) model to solve both power and computing problems in genome-wide association studies

Dr. Meng Huang[1], Dr. Zhiwu Zhang[1]
[1]*Washington State University*

Biomedical innovations have outpaced computing innovations since the completion of the human genome project. Big data accumulated from biomedical and agronomic studies, provides the potential to identify genes controlling complex human diseases and agriculturally important traits through genome-wide association studies (GWAS). However, big data also leads to extreme computational challenges, especially when complex statistical models are employed to simultaneously reduce false positives and false negatives. The newly developed Fixed and random model Circulating Probability Unification (FarmCPU) method uses a bin method under the assumption that Quantitative Traits Nucleotides (QTNs) are evenly distributed throughout the genome. The estimated QTNs are used to separate a mixed linear model into a computationally efficient fixed effect model (FEM) and a computationally expensive random effect model (REM), which are then used iteratively. To completely eliminate the computationally expensive REM, we replaced REM with FEM by using Bayesian information. To eliminate the requirement that QTNs be evenly distributed throughout the genome, we replaced the bin method with linkage disequilibrium information. The new method is called Bayesian and Linkage disequilibrium information Iteratively Nested Knitting (BLINK).

29

Evaluation of merit of including non-additive genetic variation in genomic prediction of feed-related traits in broiler chickens

Dr Yutao Li[1], Dr Rachel Hawken[2], Dr Robyn Sapp[2], Dr Andrew George[3], Dr Sigrid A Lehnert[1], Dr John M Henshall[2], Dr Antonio Reverter[1]
[1]CSIRO Agriculture, [2]Cobb-Vantress Inc., [3]CSIRO Data61

Genomic predictions of breeding value in livestock are predominately based on the detection and estimation of additive genetic effects. Non-additive genetic effects are largely ignored. With the advent of high-throughput genotyping technologies, it is feasible to derive non-additive dominance variation from marker information. In this study, we examined the consequences of including non-additive dominance variation in genomic prediction in a commercial population of 5,658 broiler chickens genotyped for 45,176 genome-wide single nucleotide polymorphism (SNP) markers. We employed mixed-model equations and restricted maximum likelihood to analyze seven feed related traits, and identified a significant proportion of the total genetic variance explained by dominance variation in all traits (ranging from 29.5% to 58.4%). Using a 5-fold cross-validation schema, we found that in spite of the large dominance component, including the estimated dominance deviations in the prediction of genomic breeding values did not improve the accuracy of the predictions for any of the phenotypes.

30

The estimation of additive, dominance and epistatic effects underlying lameness in Fleckvieh and Braunvieh cows

Tomasz Suchocki[1], Joanna Szyda[1], Magdalena Fraszczak[1], Hermann Schwarzenbacher[2], Christa Egger-Danner[2]
[1]*Wroclaw University of Environmental and Life Sciences,* [2]*ZuchtData EDV-Dienstleistungen GmbH*

The goal of the project was the statistical dissection of genetic determination of lameness in cattle. Data available for the analysis includes lameness scores, production information, pedigree and 76,934 SNP genotypes for 2981 cows representing Brownvieh and Fleckvieh breeds ascertained within the frame of the Efficient Cow project. We estimated additive and dominance effects of the SNPs and modelled the pairwise epistasis for markers selected based on their abnormal LD pattern. Mixed linear models with a fixed additive and dominance effect of SNPs and a random additive polygenic animal effect of a cow were used for this purpose. Various model selection criteria were applied in order to define the final model, which includes SNPs, showed significant additive, dominance and (or) epistatic effects on lameness. Furthermore those SNPs were functionally annotated to the UMD3.1 bovine reference genome. Gene Ontology (GO) terms and KEGG pathways underlying annotated genes were analyzed in order to identify functional clusters significantly enriched within this set of genes. Linkage disequilibrium calculation was performed at the Poznan Supercomputing and Networking Center.

31

The application of novel phenotyping technologies to reveal the genetic basis of dynamic stress-responsive traits in cotton

Dr Duke Pauli[1], Dr Pedro Andrade-Sanchez[2], Dr Greg Ziegler[3], Dr Elodie Gazave[1], Dr Andrew French[4], John Heun[2], Dr Ivan Baxter[3], Dr Tim Setter[1], Dr Kelly Thorp[4], Dr Jeffrey White[4], Dr Michael Gore[1]
[1]Cornell University, [2]University of Arizona, [3]Donald Danforth Center, USDA-ARS, [4]Arid-Land Agricultural Research Center, USDA-ARS

Heat and drought stress represent two of the most common abiotic stresses that plants encounter in modern agricultural production systems, resulting in significant economic losses. As climate change continues to increase the frequency and severity of these conditions, the development of stress-resilient cultivars becomes pivotal to sustaining crop yields. Central to meeting this challenge is the ability to elucidate the genetic and physiological basis of key stress adaptive and agronomic traits. To investigate these traits, multidimensional phenotypic data are needed that capture the dynamic response of plants to continuously changing conditions over the growing season. In light of this, we implemented high-throughput phenotyping of the plant canopy to map quantitative trait loci (QTL) controlling stress-responsive traits in a cotton population evaluated under contrasting irrigation treatments in a hot, arid environment. The ability of the field-based, mobile phenotyping system to collect data throughout the growing season revealed the temporal patterns of QTL expression in response to environmental conditions. A subset of these identified canopy trait QTL co-localized with those found to control variation for several physiological and agronomic traits, suggesting pleiotropic QTL. Furthermore, QTL for canopy temperature and normalized difference vegetative index, traits associated with leaf transpiration and wilting, co-located with candidate genes having known abiotic stress response functions. To further enhance these results, we investigated how seed ionomic profiles varied by preferential uptake of soil elements in response to drought conditions under high temperature, such as calcium which is critical for regulating stomatal aperture. This was done in combination with analyzing the ion profile of the soil itself to assess the effects of spatial variability on the observed phenotypic data. These combined results demonstrate the value of multidimensional data sets generated from novel phenotyping technologies to help provide insight into the varied physiological responses of plants to abiotic stress.

32

Studying temperature and salinity stresses in the model brown alga Ectocarpus sp. by QTL mapping

Dr. Komlan Avia[1], Dr. Susana Coelho[1], Dr. Alexandre Cormier[1], Ms. Fiona Lerck[1], Mr. Stéphane Mauger[2], Mr. Gabriel Montecinos[2], Dr. Sylvain Faugeron[2], Dr. Myriam Valero[2], Dr. Mark Cock[1], Dr. Pierre Boudry[3]
[1]Algal Genetics Group, UMR 8227 CNRS-UPMC; Station Biologique Roscoff, [2]Evolutionary Biology and Ecology of Algae, UMI EBEA 3614, CNRS, UPMC, PUCCh, UACH; Station Biologique Roscoff, [3]IFREMER - Laboratoire des Sciences de l'Environnement Marin (UMR 6539 LEMAR)

Brown algae are complex photosynthetic organisms with a very different evolutionary history than green plants. They are a dominant species in rocky coastal ecosystems. They are economically important as food products, animal feed, fertilizer, source of polysaccharides or other important extracts, or even biofuel resources. Despite their importance, many aspects of these organisms are still poorly understood. The filamentous brown alga Ectocarpus sp has been proposed as a general model for the brown algae and its genome has been sequenced.
As sessile organisms, brown algae require high levels of tolerance to various abiotic stressors (osmotic pressure, temperature, salinity, and light) and previous studies showed that an important proportion of the expressed genes are regulated in response to hyposaline, hypersaline or oxidative stresses.
Using double digest RAD sequencing method, a denser genetic map has been constructed and QTL for two stress-related phenotypes (tolerance to high temperature and low salinity) have been mapped. These first results of QTL mapping in Ectocarpales are discussed in the light of the peculiar as well as common characteristics of these species relative to other algae and land plants.

33

Accuracy of genomic breeding values for meat tenderness in Polled Nellore cattle

DR. CLAUDIO U MAGNABOSCO[1], DR FERNANDO B LOPES[1,2], DR BRUNO VALENTE[2], DR EDUARDO C EIFERT[1], DR GUILHERME J ROSA[2], DRA LUCIANA A REGITANO[1], DR MARCOS F COSTA[1], DR ROBERTO D SAINZ[3]

[2]Brazilian Agricultural Research Corporation (EMBRAPA), [2]University of Wisconsin-Madison, [3]University of California-Davis

Nellore cattle comprise more than 80% of the beef cattle in Brazil given their adaptability, rusticity, and high resistance to ectoparasites. Despite their advantages for production in tropical environments, Zebu cattle in general lack satisfactory tenderness in their meat. Although meat tenderness has an important genetic component underlying it, genetic improvement programs using traditional selection approaches is hampered by constraints on phenotypic evaluation of meat quality. Therefore, genomic selection may be a good strategy to improve meat quality traits. This study aimed to compare the accuracies of different Bayesian regression models in predicting molecular breeding values for meat tenderness in Polled Nellore cattle. The dataset was composed of 427 animals slaughtered between 2005 and 2012. Meat tenderness was measured by Warner-Bratzler Shear Force (WBSF) of the Longissimus muscle. Ninety animals were genotyped in Illumina BovineHD (777K) and 337 animals in GeneSeek Genomic Profiler (77K). The FImpute program was used for imputation. The quality controls of SNP were HWE p-value≥0.1%; MAF>1% and Call Rate>90%. The effect of each SNP was estimated using Ridge Regression, Lasso, BayesA, Bayes B and Bayes Cπ methods. The prediction accuracy was assessed by the correlation between genomic breeding value and the observed meat tenderness phenotype. The prediction accuracies obtained by Bayesian Ridge Regression, Bayesian Lasso, Bayes A, Bayes B and Bayes Cπ models using all markers were all very similar, ranging from 0.22 (Bayes Cπ) to 0.25 (Bayes B). Some small advantage was observed with the Bayes B approach when higher numbers of markers were pre-selected based on their p-values resulting from a GWAS analysis. Although this research used relatively few animals, the design of the segregating population ensured wide genetic variability for meat tenderness, which was important to achieve good results of accuracy of genomic selection.
Keywords: Bayesian regression models, genomic prediction, Nellore, meat tenderness

34

Bioeconomic selection index (EMGTe) for Nellore Brazil breeding program

Dr Raysildo Lobo[1], Dr Fernando Baldi[2], Dr Luis Gustavo Figueredo[3], Dr Henrique Nunes de Oliveira[2], Dr Cláudio de U. Magnabosco[4], MSc Luiz Antonio F. Bezerra[1], Dr Carina U. Faria[5]
[1]USP, [2]UNESP FCAV, [3]ANCP, [4]EMBRAPA CERRRADOS, [5]UFU

The aim was to develop a bioeconomic model in order to develop a selection index for Nellore Brazil breeding program. The bioeconomic model simulates the biological and economic components of a commercial Nellore beef cattle system, based on the central region of Brazil (Cerrado biome), which is the region with the greatest meat production in Brazil. The simulations were performed using the Excel 2007 spreadsheet, Microsoft Office programs. The information used was based on real and current market data as well as various scientific studies available in the literature. The economic values for the following traits were calculated: direct (W) and maternal (MW) weight at 120 (W120 and MW120), 210 (W210 and WP210), 365 and 450 days of age (W365 and W450), ribeye area (REA) and fat thickness (BFT) obtained by ultrasonography, scrotal circumference at 365 and 450 days of age (SC365 and SC450), age at first calving (AFC), probability of early calving (3P), stayability (STAY) and total cow accumulated productivity (PAC). To estimate the relative economic weighted for each trait, the economic values obtained in bioeconomic model and genetic parameters, heritability and genetic correlations between the traits, were considered. The genetic parameters were estimated from the Nellore Brazil breeding program database. The bioeconomic selection index of the Nellore Brazil breeding program was developed using the methodology of Hazel (1943). After several simulations carried out with several combinations of traits, the definitive bioeconomic selection index, called EMGTe, was developed. The following traits with their respective relative economic weights were considered in the EMGTe: STAY (22%), W450 (24%), W210 (16%), REA (9%), 3P (9%), AFC (6%), MW210 (5%) SC365 (3%), SC450 (3%) and MW120 (3%). The EMGTe would be utilized by the breeders to rank the candidates to selection using the genomic breeding values predicted in the Nellore Brazil breeding program.

35

Exploring the last chromosome: Y-linked sequence variation in the house mouse

Mr. Andrew Morgan[1], Fernando Pardo-Manuel de Villena[1]
[1]*University of North Carolina*

Over 180 million years since their divergence from autosomes, mammalian Y chromosomes have evolved a gene repertoire highly specialized for male reproduction. Y chromosomes are also a valuable tool for population genetics and phylogeny because they are inherited without recombination from a single parent. The Y chromosome of the house mouse (Mus musculus) is exceptional in that it is almost entirely euchromatic and has important roles in speciation. However, its complex repetitive structure has hindered analyses by high-throughput sequencing. We sought to fill this gap by performing a systematic survey of Y-linked variation in mouse. We have compiled whole-genome sequencing data for 132 male samples including wild-caught and laboratory representatives of each of the three subspecies of the house mouse (Mus musculus domesticus, M. m. musculus and M. m. castaneus) and transcriptome data from outgroup specices. First, we show that both the short and long arms of the Y chromosome of the three subspecies are differentiated by megabase-sized structural variants. Next we apply several complementary approaches to obtain a catalog of sequence variants in the non-ampliconic portion of the Y chromosome. Using laboratory strains with known pedigree relationships, we estimate the average mutation rate on Y to be $9.1 \pm 1.8 \times 10^{-9}$ per site per generation. Finally we use approximate Bayesian computation to compare patrilineal and matrilineal demographic patterns in wild mice. We find evidence for increased geographic differentiation and reduced effective population size on the Y compared to the mitochondria, and attribute this to the combined effects of background selection and variability in male reproductive success.

36

Modeling genome by environment interaction in allohexaploid wheat (Triticum aestivum)

Nicholas Santantonio[1], Dr. Mark Sorrells[1], Dr. Jean-Luc Jannink[1]
[1]*Cornell University*

As an autogamous allohexaploid, wheat (Triticum aestivum) represents a unique opportunity to study the effects of multiple genomes, A, B and D, as well as their interactions with each other and their environment. The joining of orthologous genes during an allopolyploid event results in homeoalleles between genomes that can have additive effects as well as non-additive-interactions. The relative importance of each genome may also depend on the environment in which it is grown, as well as the presence of genome by genome interactions. To investigate the effect of genome by environment and genome by genome interactions, genome specific additive genetic covariance matrices were calculated from markers belonging to each genome, respectively. Using these genome specific covariance matrices, genome specific additive effects and variances can be estimated using REML in a linear mixed model with separate kernels for each genome. Similarly, additive by additive between genomes and genome by environment interaction terms can be included in the model. This methodology is applied to three multi-environment trial data sets of winter and spring wheat grain yield to determine the relative importance of genomes across environments. The change in the proportion of genetic variance contributed by each genome suggests that the relative importance of genome is environment specific. Similarly, differences in the proportion of variance of each genome across environments varied between populations. Genome by genome interaction variances were mostly zero, and in a few cases resulted in a failure of REML convergence, and are therefore not shown. The non-independence of genome specific covariance matrices due to non-random mating may result in somewhat erroneous results, particularly in small populations and more research is needed to address this issue. This methodology may have additional uses for specific breeding goals, such targeted selection of individual genomes for specific environments.

38

Heritability of mentality traits in Swedish dogs

Dr Erling Strandberg[1], Dr Katja Nilsson[1], Dr Kenth Svartberg[2]
[1]*Swedish Univ of Agricultural Sciences,* [2]*Svartbergs Hundkunskap*

The dog mentality assessment (DMA) has been used in Sweden since 1989 to describe mentality in dogs. DMA was developed mainly for working dogs and there was a demand for a mentality assessment suitable also for other breeds. With this aim, the new behavior and personality assessment in dogs (BPH) was created and assessments were started in 2012. The aim of this study was to estimate heritabilities (h^2) for mentality traits for 5 breeds: Rhodesian Ridgeback (RR, n=493), Labrador Retriever (LR, n=491), Nova Scotia Duck Tolling Retriever (NSDTR, n=425), Staffordshire Bullterrier (SBT, n=343), and American Staffordshire Terrier (AST, n=298). From the over 250 markings on the score sheet, 6 summarizing mentality traits were defined across the 5 breeds using exploratory factor analysis. Average h^2 for all traits within breed varied from 0.29-0.38. Hostility had the highest h^2 across breeds but also the largest variation (0.38, range 0.18-0.62). Confidence had more consistently high h^2 (0.36, range 0.22-0.45). The other four traits (Playfulness, Sociability, Curious and Confident, and Positive and Energetic) had average h^2 of around 0.3, with range from 0.10 to 0.59. RR had the highest average heritabilities and the estimates varied the least (0.38, SD 0.03), whereas NSDTR had the lowest heritabilities (0.29, SD 0.06). Standard errors ranged from around 0.10 for the largest breeds to about 0.15 for the smallest. These results show that the traits defined in the new BPH have at least as high h^2 as the traits in DMA. Continued work will focus on estimation of genetic correlations between, on one hand, BPH-traits, and both DMA-traits and everyday behavior traits based on owner questionnaire data, on the other hand.

39

Genetic mapping by bulk segregant analysis in Drosphila: Experimental design and simulation-based inference

John E Pool[1]
[1]*University of Wisconsin - Madison*

Identifying the genomic regions that underly polygenic phenotypic variation is a key challenge in modern biology. Many approaches to quantitative trait locus mapping in animal and plant species suffer from limited power and genomic resolution. Here, I investigate whether bulk segregant analysis (BSA), which has been successfully applied for yeast, may have utility for trait mapping in Drosophila (and other organisms that can be experimentally bred in similar numbers). I performed simulations to investigate the statistical signal of a quantitative trait locus (QTL) in a wide range of BSA and introgression mapping (IM) experiments. Results suggested that BSA consistently provides more accurate mapping signals than IM (in addition to allowing the mapping of multiple traits from the same experimental population). The performance of BSA was maximized by having multiple independent crosses, more generations of interbreeding, larger numbers of breeding individuals, and greater genotyping effort, but was less affected by the proportion of individuals selected for phenotypic extreme pools. I also introduce a prototype analysis method of Simulation-based Inference for BSA Mapping (SIBSAM). In addition to identifying significant QTLs and estimating their genomic confidence intervals and relative effect sizes, this method tests whether overlapping peaks should be considered as two distinct QTLs. This approach will facilitate improved trait mapping in Drosophila and other species for which hundreds or thousands (but not millions) of offspring can be generated and analyzed.

40

Rare variants from imputed whole genome sequence explain a small fraction of the total genetic variance in different traits in cattle

Mrs. Qianqian Zhang[1,2], Dr. Bernt Guldbrandtsen[1], Prof. Mogens Sandø Lund[1], Dr. Goutam Sahana[1]
[1]*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University,*
[2]*Animal Breeding and Genomics Centre, Wageningen UR Livestock Research*

Theoretical and empirical studies suggest that rare variants (defined as those have minor allele frequency MAF < 0.01), could play a significant role in quantitative trait variation. We studied the contribution of rare genetic variants to the total genetic variance for two traits in cattle, stature (h^2=0.60) and milk yield (h^2=0.35). The response variable used was de-regressed proofs (DRPs) from ~5000 Nordic Holstein bulls. The number of genetic variants in the imputed whole genome sequence after quality control was ~15 million. Two multicomponent models were analyzed where stratifications were based on MAF and region-specific linkage disequilibrium (GREML-MS and GREML-LDMS) as implemented in the GCTA software. The total variances of DRPs of stature and milk yield explained by all sequence variants were 84.5% and 87.4% respectively. The SNPs with MAF < 0.01 accounted for 2.67% and 1.56%; in comparison, common SNP variants (MAF>0.01) explained 81.8% and 85.8% of variance for stature and milk yield DRPs respectively. These results suggest that rare variants explain a small percentage of the total genetic variance for stature and milk yield in cattle.

41

Joint genetic evaluation of mastitis resistance and recovery ability in danish dairy cows under automated milking systems

Mr B.G. Welderufael[1,2], L.L.G. Janss[2], D.J. de Koning[1], L.P. Sørensen[2], P. Løvendahl[2], W.F. Fikse[1]
[1]*Swedish University of Agricultural Sciences,* [2]*Aarhus University*

Mastitis in dairy cows is an unavoidable problem and variation in recovery from mastitis is therefore of interest, in addition to resistance to mastitis. Genetic parameters for mastitis resistance and recovery were estimated for Danish Holstein-Friesian cows using data from Automatic Milking Systems equipped with online somatic cell count (OCC) measuring units. The OCC measurements were converted to elevated mastitis risks (EMR), a continuous variable (on a [0-1] scale) indicating the risk of mastitis. EMR values above 0.6 were assumed to indicate that a cow had mastitis. For each cow and lactation the sequence of health states (mastitic or healthy) was converted to weekly transitions: 0 if the cow stayed within the same state and 1 if the cow changed state. The result was two series of transitions: one for healthy to diseased (HD, to model mastitis resistance) and the other for diseased to healthy (DH, to model recovery). The two series of transitions were analyzed with a bi-variate threshold model including a linear combination of systematic effects and of a function of time. The model included effects of herd and parity, herd-test-week, permanent environment (to account for the repetitive nature of transition records from a cow) plus two time-varying effects (lactation stage and time within episode). In early lactation, there was an increased risk of getting mastitis but the risk remained stable afterwards. Mean recovery rate was 45.7%. Posterior mean of correlations between the traits were -0.853 (PSD=0.108), -0.909 (PSD=0.076), -0.292 (PSD=0.105) and 0.052 (PSD=0.157) for the random effects of genetic, permanent environment, cow-parity interaction, and herd-test-week, respectively. Heritabilities were 0.080 (PSD=0.026) for HD and 0.075 (PSD=0.030) for DH. Recovery could be considered as a new trait for selection even though the negative correlation suggests that a cow that easily contracts mastitis has greater difficulty to recover.

44

Proportion of inter-individual differences for time to death of glioblastoma multiforme incorporating multi-omic layers.

Yeni Liliana Bernal Rubio[1], Hank Wu[1], Corinne Griguer[2], Juan Pedro Steibel[1], Agustin Gonzalez[1], Ana Ines Vazquez[1]
[1]*Michigan State University,* [2]*University of Alabama at Birmingham*

Glioblastoma multiforme (GBM) has been recognized as the most common type of malignant brain tumor. This tumor type grows rapidly, spreading into brain tissue and becoming highly fatal. The goal of this work was to estimate the proportion of inter-individual differences for survival time associated to SNP and Gene Expression (GE) in GBM patients (n = 295). Dataset is made available by The Cancer Genome Atlas, and includes clinical covariates (CLIN) such as race, gender and method of diagnosis among others, gene expression on the tumor (GE, 10,848 genes) and SNP genotypes from normal tissue (509,197 SNP). Three G-BLUP models were implemented with the BGLR package: (a) A model including CLIN and GE; (b) A model including CLIN and SNP; (c) A model including CLIN, SNP and GE. Time to death was considered as response variable right censored. The sum of GE effects ($\mathbf{u}_{ge}$) was incorporated into the model as random effect, which is assumed to follow a Normal distribution $\mathbf{u}_{ge} \sim N(\mathbf{0}, \mathbf{G}_{ge}\sigma^2_{u\_ge})$, where $\mathbf{G}_{ge}$ is a GE-derived similarity matrix between GBM tumors. SNP effects of all subjects ($\mathbf{u}_{snp}$) were assumed to be Normally distributed $\mathbf{u}_{snp} \sim N(\mathbf{0}, \mathbf{G}_{snp}\sigma^2_{u\_snp})$, where $\mathbf{G}_{snp}$ is the SNP-derived genomic relationship matrix between patients. In all cases, a residual term, $\boldsymbol{e} \sim N(\mathbf{0}, \boldsymbol{I}\sigma^2_e)$,was included. The ratio of variances $\sigma^2_{u\_\mathrm{ge}}/(\sigma^2_{u\_\mathrm{ge}} + \sigma^2_e)$, $\sigma^2_{u_snp}/(\sigma^2_{u_snp} + \sigma^2_e)$and $(\sigma^2_{u\_snp} + \sigma^2_{u\_\mathrm{ge}})/(\sigma^2_{u\_snp} + \sigma^2_{u\_\mathrm{ge}} + \sigma^2_e)$ were obtained for CLIN+GE, CLIN+SNP and CLIN+SNP+GE respectively. These ratios were 0.254, 0.453 and 0.474 respectively. In the last case, variances $\sigma^2_{u\_snp}$ and $\sigma^2_{u\_\mathrm{ge}}$ were 0.318 and 0.233 respectively. Interestingly, $\sigma^2_{u\_\mathrm{ge}}$ was smaller than $\sigma^2_{u\_snp}$, probably because the GE in the tumor is relatively homogeneous compared to other cancers. Therefore, integration of whole-genome omics as GE and SNP allows to recover a substantial proportion of variance, constituting an important tool in the analysis of factors affecting the progress of GBM and potentially other cancers too.

45

Genome-wide association analysis in dogs implicates 98 loci as risk variants for cranial cruciate ligament rupture

Lauren Baker[1], Brian Kirkpatrick[1], Guilherme J. M. Rosa[1], Daniel Gianola[1], Bruno Valente, Julia Sumner[2], Wendy Baltzer[3], Zhengling Hao[1], Emily Binversie[1], Nicola Volstad[1], Susannah J. Sample[1], Peter Muir[1]
[1]University Of Wisconsin-madison, [2]Louisiana State University , [3]Oregon State University

Anterior cruciate ligament (ACL) rupture is a common condition that can be devastating and life changing, particularly in young adults. Contralateral ACL rupture or rupture of ACL graft repair is also common. Cruciate ligament rupture is also common in dogs and is an excellent naturally occurring model for human ACL rupture. Disease prevalence exceeds 5% in several breeds, ~100-fold higher than humans. We provide insight into the genetic etiology of the human ACL rupture condition by performing a genome-wide association study (GWAS) in dogs using three linear mixed models. A Bayesian mixture model was also used to determine genomic architecture of the trait. Heritability was also estimated. In the dog, GWAS of 98 cases and 139 controls from the Labrador Retriever breed identified 129 single nucleotide polymorphisms (SNPs) that reside in 98 risk loci. Associated regions (P<5.0E-04) explained approximately half of the phenotypic variance in the ACL rupture trait. Five of these regions were located in uncharacterized or non-coding regions of the genome. A SNP on chromosome 24 that resides in a 5kb haplotype block met genome-wide significance (P=3.63E-06). Genes within this block include BPI, LBP, RALGAPB, ADIG, ARHGAP40, SLC32A1, ACTR5, PPP1R16B, FAM83D, and DHX35. Genotypic risk estimated for each dog based on the risk contributed by each GWAS locus showed clear separation of ACL rupture cases and controls. GWAS pathways were enriched for aggrecan signaling and a gene set encoding membrane transport proteins with a variety of physiological functions. Bayesian analysis indicated the majority of risk SNPs have small to moderate effects on genetic variation. Broad and narrow sense heritability was estimated at 0.476 and 0.513 respectively. In summary, this study provides genetic evidence that ACL rupture is a moderately heritable highly polygenic complex trait. Our results implicate aggrecan signaling pathways in the ACL rupture condition in dogs.

46

Community genomics of plant-insect interactions

Mrs. Hilary L. Barker[1], Dr. Jennifer F. Riehl[1], Dr. Liza Holeski[2], Dr. Pär Ingvarsson[3], Dr. Richard L. Lindroth[1]
[1]*UW-Madison,* [2]*Northern Arizona University,* [3]*Umeå University*

Community genomics seeks to understand how intraspecific genetic variation in one species structures communities of interacting species. Within this framework, research has shown that different genotypes of foundation tree species (e.g., Populus) harbor analogously different insect communities. Insect communities can then be considered as a heritable trait of the tree. Yet, the particular Populus genes underlying this relationship have not been identified. We have established a community genomics resource, the Wisconsin Aspen ("WisAsp") genetic mapping garden. The garden contains 515 replicated genotypes of aspen (P. tremuloides) that were collected from throughout Wisconsin and planted in 2010. We surveyed both tree traits (e.g., leaf phenology, foliar chemistry) and insect communities in 2014-15. We also sequenced 60,000 SNP markers (1-2 probes/gene) for genome wide association (GWA) mapping. Thus far we have found substantial genotypic variation in tree traits and insect communities (e.g., foliar defense compounds vary by 24-fold), and have found evidence that the tree traits influence the composition of the insect communities (e.g., timing of bud break is correlated with insect community nonmetric multidimensional scaling [NMDS] coordinates, $P < 0.001$). Here we report the initial findings of the GWA results that link Populus genes to tree traits and insect communities. This work provides the mechanistic link between intraspecific genetic variation in a foundation species and effects on communities of interacting organisms. Concepts from this work can be applied in conservation initiatives (e.g., protecting plant populations with a genetic makeup that promotes biodiversity in associated communities) and crop management (e.g., selecting crop varieties with genotypes that attract beneficial insect communities rather than focusing on resistance to one insect pest species). Ultimately, this research will also transform our understanding of evolution, since once we identify the genetic underpinnings of a community, we can then search for signs of community evolution.

47

Bayesian networks illustrate phenotypic and genomic trait connections in maize

Katrin Töpner[1], Guilherme J.M. Rosa[2], Daniel Gianola[2], Chris-Carolin Schön[1]
[1]*Plant Breeding, Technical University of Munich,* [2]*Animal Sciences, University of Wisconsin-Madison*

Relationships between phenotypic traits in maize were investigated using a new methodology and data representing the European heterotic pools Dent and Flint. This included inference of genomic and phenotypic relationships by Bayesian Networks (BN), and an assessment of the networks with Structural Equation Models (SEM). The methodology involved three steps. First, a Bayesian multiple-trait Gaussian model (MTM) was applied to decompose phenotypic values into their genomic and residual components. Second, genomic and phenotypic network structures among traits were learned from these components. Network learning was realized with six different algorithms for comparison, two score-based and four constraint-based approaches. Third, SEM analyses ranked the networks and thereby the algorithmic settings in terms of model fit and predictive ability, and compared these with the standard assumption of a full network among traits. Features of a suitable algorithm for this context could be identified. Inferences on genomic and phenotypic trait connections were depicted in directed acyclic graphs. These illustrate conditional independence and potential causal links among traits and provide information beyond mere pairwise association between traits. In fact, some high genomic correlation could be shown to be spurious. Phenotypic networks were alike between Dent and Flint as expected whereas genomic networks differed. By reference to a recent QTL and GWA study with our data, we also highlighted possible impact of our findings on marker effect estimation and association testing.

48

Sequence-level partitioning of genomic variance for mastitis and production trains in dairy cattle based on transcrptomic expression data

Mr. Lingzhao Fang[1], Dr Goutam Sahana[1], Dr Guosheng Su[1], Dr Ying Yu[2], Dr Shengli Zhang[2], Dr Mogens Sandø Lund[1], Dr Peter Sørensen[1]

[1]*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University,* [2]*Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture & National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University*

Transcriptomics profiling studies in dairy cattle have shown that the genome regions differentially expressed during intra-mammary infection (IMI) with mastitis-causing pathogens are enriched for the genes functionally involved in innate immune responses and lipid/protein metabolism. We wanted to investigate if transcriptionally active genome regions during IMI are enriched for genetic markers associated with mastitis and milk production traits in dairy cattle. Genome regions were identified from several IMI studies involving three tissues (blood, liver and udder) and three pathogens (E. coli, E. coli endotoxin (LPS) and S. aureus). Single marker test statistics from genome-wide association analyses of mastitis and production traits in Nordic Holstein, Nordic Red Cattle and Danish Jersey were obtained using imputed sequence-level SNPs. Enrichment of associated genetic markers in transcriptionally active regions was assessed by comparing the sum of the squared single marker test statistics within these regions to an empirically derived distribution under an competitive null hypothesis. We found that the genes whose expression levels were highly upregulated in liver during IMI were enriched for functional processes involved in inflammatory responses and were enriched for genetic markers associated with mastitis. In contrast, the genes highly down-regulated in liver during IMI were involved in multiple biological processes including fatty acid metabolism and were enriched for genetic markers associated with milk production traits. Furthermore, our results indicated that in the genome regions which were transcriptionally active during IMI the average genomic variance explained by each SNP was higher for both mastitis and production traits compared to the remaining genome regions in all three dairy cattle breeds.

49

Prediction for candidates two or more generations away from the reference population

Dr. Peipei Ma[1], Professor Mogens S. Lund[1], Dr. Guosheng Su[1]
[1]*Aarhus University*

Previous studies have showed the prediction accuracy decreased when the relationship between reference and test individuals is distant. Due to heavy use of genomic selection, most selection candidates are now two or three generations away from the nearest ancestors in the reference population. The aim of this study is to investigate strategies to increase prediction reliability for these candidates. Eurogenomics data were used in this study, including milk yield, fat yield, protein yield, fertility and mastitis. Deregressed proofs (DRP) were used as response variables for genomic prediction. To validate the genomic predictions, the Danish bulls born between 2006 and 2010 were used as test animals, and individuals born before 2006 were used as the reference set (Full). To mimic the scenarios that the candidates are two or three generations away from the reference, individuals from reference with a pedigree kinship larger than 0.4 (Two_gen) or 0.2 (Three_gen) with any of the candidates were removed from the Full reference set. To keep the size of reference sets constant, individuals in Full and Two_gen were reduced to reach the same number of individuals as Three_gen, by choosing a subset randomly or a subset highly related to the candidates. Genomic best linear unbiased prediction (GBLUP), GBLUP with a residual polygenic effect and Bayesian Mixture models were used. The results confirmed that the prediction reliability using all the models decreased when the relationship between reference and test set decreased. The reliabilities decreased less for Bayesian methods compared with GBLUP method. The prediction reliability when including a polygenic effect in the model became lower when the candidates were far from the reference individuals compared with GBLUP without a residual polygenetic effect.

50

Assessment of bias due to genomic pre-selection in the Dutch-Flemish breeding value estimation

Dr. Herwin Eding[1], Davis Krezins[2]
[1]*CRV,* [2]*Latvia University of Agriculture*

More and more countries have implemented genomic evaluation and selection in dairy cattle. Genomic selection based on massive marker information can increase accuracy of selection of breeding animals significantly. However, due to intensive pre-selection of young bulls, bull sire EBVs are likely biased. A quick fix to eliminate pre-selection bias is removing one or both parents of pre-selected sire from the pedigree. We applied this quick fix to assess the magnitude of pre-selection bias in the Dutch/Flemish dairy genetic evaluation for the production of milk, fat and protein. This was done by comparing breeding values of bull sires obtained from an evaluation with a full pedigree to those from an evaluation with a pruned pedigree. We investigated 1) to what extent the potential for pre-selection bias is, as well as 2) the actual bias in the current estimate of GEBV. The results show an amount of potential bias that was 5%, 4% and 3.5% of the genetic standard deviation for milk, fat and protein, respectively for bulls with between 10 and 50 daughters. The bias was downward as a result of elevated means in the group of pre-selected progeny. This bias, however, vanished quickly to less than 0.1% when number of daughters increased. Hence, pre-selection bias seems a temporary issue for young bulls only. The bias estimates from the current Dutch/Flemish evaluation never exceeded 1% of genetic standard deviation. Moreover, these estimates did not show a consistent sign. Hence, bias due to pre-selection currently does not play a significant role in the breeding value estimation in the Netherlands.

51

Genomic prediction accuracies in commercial barley and wheat schemes

Dr. Vahid Edriss[1], Dr. Fabio Cericola[2], Dr. Jens D Jensen[1], Dr. Jihad Orabi[1], Dr. Jeppe R Andersen[1], Prof. Just Jensen[2], Dr. Ahmed Jahoor[1]
[1]Nordic Seed, [2]Aarhus University

Genomic prediction is a novel method applied in plant breeding to identify the individuals with highest breeding values to contribute in the next generations. It uses the SNP markers spread over the whole genome. There are two main advantages in using genomic prediction: 1) it can predict individual merits on early stage of growth that can shorten each cycle in the breeding program; 2) reduction in cost and laboring with less amount of phenotyping. The aim of this study was to evaluate the benefit of genomic prediction in spring barley and winter wheat breeding programs. The lines were phenotype from 2013 to 2015. Each year 300 new lines were phenotyped at three locations with three replicates and in the year after same 300 lines were also phenotype in two of the locations with three replicates for yield and agronomic traits. A total of 900 advanced lines from each of the two crops were genotyped with iSelect Illumina BeadChips of 9K for barely and 15K for wheat. Genomic prediction was carried out using GBLUP model. The model consisted of fixed effect of year×location×trial and random effects of genotype, genotype×location plus residual. The prediction accuracy was calculated as the correlation between corrected phenotype for fixed effect and genomic predicted breeding values. The results showed moderate to high prediction accuracies for agronomic and quality traits in both barley and wheat. The prediction accuracies were yield = 0.53, Plant height = 0.66, heading date = 0.68 and seed protein percentage = 0.54 in barely. In wheat accuracies were yield = 0.52, lodging = 0.56, seed protein percentage = 0.61 and seed starch = 0.57. The overall conclusion from the results is that genomic prediction has the potential to be incorporated into commercial breeding programs.

52

Mapping variants to gene ontology categories improves genomic prediction for quantitative traits in Drosophila melanogaster

Dr. Peter Sørensen[1], Dr. Stefan M Edwards[2], Dr. Izel F Sørensen[1], Dr. Pernille Sarup[1], Professor Trudy FC Mackay[3]

[1]Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, [2]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, [3]Department of Biological Sciences and Program in Genetics, North Carolina State University

Predicting individual quantitative trait phenotypes from high resolution genomic polymorphism data is important for personalized medicine in humans, plant and animal breeding and adaptive evolution. However, this is difficult for populations of unrelated individuals when the number of causal variants is low relative to the total number of polymorphisms and causal variants individually have small effects on the traits. We hypothesized that mapping molecular polymorphisms to genomic features such as genes and their gene ontology categories could increase the accuracy of genomic prediction models. We developed a genomic feature best linear unbiased prediction (GFBLUP) model that implements this strategy and applied it to three quantitative traits (startle response, starvation resistance and chill coma recovery) in the unrelated, sequenced inbred lines of the Drosophila melanogaster Genetic Reference Panel. Our results indicate that subsetting markers based on genomic features increases the predictive ability relative to the standard GBLUP model. Both models use all markers, but GFBLUP allows differential weighting of the individual genetic marker relationships, whereas GBLUP weighs the genetic marker relationships equally. Simulation studies show that it is possible to further increase the accuracy of genomic prediction for complex traits using this model, provided the genomic features are enriched for causal variants. Our GFBLUP model using prior information on genomic features enriched for causal variants can increase the accuracy of genomic predictions in populations of unrelated individuals and provides a formal statistical framework for leveraging and evaluating information across multiple experimental studies to provide novel insights into the genetic architecture of complex traits.

53

Co-expression gene networks reveal potential biomarkers for reduction of boar taint in pigs

Mr Markus Drag[1], Dr Ruta Skinkyte-Juskiene[1], Dr Duy N. Do[2], Dr Lisette J.A. Kogelman[3], Professor Haja N. Kadarmideen[1]

[1]University Of Copenhagen, Department of Large Animal Sciences, [2]McGill University, Department of Animal Science, , [3]Danish Headache Center & Department of Neurology, Rigshospitalet Glostrup,

Boar taint (BT) is an offensive odour or taste of porcine meat which may occur in entire male pigs due to skatole and androstenone accumulation. To avoid BT, castration of young piglets is performed but this strategy is under debate due to animal welfare concerns. The study aimed to reveal potential BT biomarkers for optimized breeding. Male pigs (n=48) with low, medium and high genetic merit of BT were selected and tissues from liver and testis were subjected to transcriptomic profiling by RNA-Seq. The reads were mapped to the Sus scrofa reference genome (Ensembl, ver. 79) which resulted in ~87% uniquely mapped reads. Quality control by Qualimap revealed ~51% of reads mapped in the exonic. Differential expression (DE) comparison of low, medium and high BT using Limma revealed a 10-fold difference in numbers of DE genes (FDR < 0.05) between liver and testis, with testis being the highly active tissue. GOseq was used to find enriched gene ontology (GO) terms and REVIGO was used to filter semantic similarities. In both liver and testis, a GO termed “oxidoreductase activity” was enriched ($p < 0.05$) due to high amounts of 5α-reductases involved in steroid metabolism, including androstenone synthesis. In testis, >80 DE genes were functionally classified by the PANTHER tool to “Gonadotropin releasing hormone receptor” and “Wnt signaling” pathways which play a role in reproductive maturation and proliferation of spermatogonia, respectively. WGCNA was used to build co-expression networks and enrichment analysis and semantic filtering revealed the GO terms “catalytic activity” and “transferase activity” to be overrepresented ($p < 0.05$) in liver and testis, respectively. Transferases include prenyltransferases which are involved in catalysis of the precursor of steroid hormones. Extraction of hub genes from important modules and integration with DE results revealed potential biomarkers for BT.

54

Polymorphisms in MCT1 and CD147 genes in Arabian and Quarter Horses

Ms Inaê C. Regatieri[1], Dr. Guilherme L. Pereira[2], Dr. Rogério A. Curi[2], Ms Maria Luiza M. Almeida[1], Dr. Guilherme C. Ferraz[1], Dr. Antonio Queiroz-Neto[1]
[1]FCAV/UNESP, [2]FMVZ/UNESP

Monocarboxylate transporter isoform 1 (MCT1) and its ancillary protein, CD147, are expressed In muscle and erythrocyte membranes. They function to transport H+ and lactate ions from the plasma into red blood cells, maintaining acid/base balance and retarding systemic acidosis and muscular fatigue. The study of polymorphisms in MCT1 and CD147 genes and their affects in the protein formation and transport of lactate and H+ ions, can provide insight for genetic technologies used for breeding selection of athletic equines. This study investigated polymorphisms of MCT1 and CD147 genes in Arabian and Quarter Horses in order to associate polymorphisms with performance. Arabian horses were divided into high and low performance groups based on success in endurance competition, whereas racing Quarter Horses were separated by a speed index. Polymorphisms of the MCT1 and CD147 genes were analysed by sequencing. To compare the frequencies of polymorphic single nucleotide polymorphisms (SNPs) between the groups, the Fisher's exact test was performed in the software R with significance level of 5%. The A allele from the polymorphisms Lys457Gln:1573A>C of MCT1 and Ile51Val:168A>G of CD147 gene were essentially fixed in both breeds. Only two Arabian horses from the high performance and one from the low performance group appeared heterozygous for the AC polymorphism in MCT1. In summary, the A alleles, corresponding to the polymorphisms in MCT1 and CD147 genes, have shown to be fixed or nearly fixed in Arabian and Quarter Horses. It was not possible to determine the influence of the polymorphisms in MCT1 and CD147 genes in the athletic performance of Arabian and Quarter Horses. It is likely that with a higher number of animals in future studies, some association can be found between the polymorphism and performance traits in athletic horses.

Keywords: equine; fatigue; lactate; performance; sequencing

55

Accuracies of genomic prediction of breeding values for growth, carcass and meat quality traits in a crossbred sheep population including non-additive effects

Luiz F Brito[1], Shannon M Clarke[2], John C McEwan[2], Natalie Pickering[3], Stephen P Miller[2], Wendy Bain[2], Flávio S Schenkel[1]
[1]University of Guelph, [2]AgResearch, [3]Focus Genetics

The New Zealand sheep industry has been characterized by a high and increasing proportion of composite breeds and crossbred animals. The genetic evaluation models currently used focus on additive genetic effects. However, other genetic effects, such as dominance, may also play an important role on the performance of composite and crossbred animals. The objectives of this study were: 1) to estimate genomic additive and dominance variances, and 2) to compare the accuracy of prediction of an additive genetic effect model to an additive and dominance effects model, using a large dataset and several traits with different genetic architecture (n = 47). A total of 14,634 individuals born between 2010 and 2014, genotyped for the High-Density Ovine BeadChip (606,006 SNP) and measured for various growth, carcass and meat quality traits (e.g. weaning weight, carcass weight, dressing percentage, butt circumference, eye muscle depth, tenderness, meat colour and colour stability, etc.) were included in the analyses. The animals were primarily progeny from terminal sire composites (i.e. Primera and Lamb Supreme) and Texel mated to a variety of maternal/dual purpose breeds. The software GVCBLUP was used to estimate variance components and to predict mBVs. Trait phenotypes adjusted for fixed effects were used as the response variables. For some traits, dominance effects account for a substantial proportion of the total genetic variance with a range from zero to 65% (average: 18% ± 14%). The differences in accuracies between the additive and dominance effects model and the additive effect model ranged from -15.8% to 16.6%, with an average of 0.2%. Only three traits out of 47 showed an increase in accuracies higher than 2% (i.e. meat colour traits and weaning weight) when including dominance effects. Therefore, with exception of these traits, there were no clear benefits of including dominance effects in the model to predict mBVs.

56

Bias due to selective genotyping in genomic prediction using H-BLUP

Lei Wang[1], Per Madsen[1], Robyn Sapp[2], John Henshall[2], Rachel Hawken[2], Setegn Worku Alemu[1], Anders Christian Sørensen[1], Just Jensen
[1]*Aarhus University,* [2]*Cobb-Vantress Inc.*

H-BLUP uses a variance-covariance structure based on a combined relationship matrix (H), which augments a pedigree-based relationship matrix (A) with a genomic relationship matrix (G) for genotyped individuals. In practice, often only preselected individuals are genotyped and this selective genotyping may generate bias in estimated variance components (VC) and predicted breeding values (EBVs). Most current applications of H-BLUP use VC from traditional animal models, and use observed allele frequencies computed from genotyped individuals only to construct G, since base population allele frequencies are usually not available. However, such VC doesn't take ongoing selection based on genomic information into account. The aim of this study is to investigate the effects of selective genotyping and of different strategies to compute G on the accuracy and potential bias in VC estimates by using H-matrix, and in EBVs from H-BLUP. We simulated a population similar to a selection line of broilers with overlapping selection rounds. In the simulation, there were 20 rounds of BLUP selection after base population, followed by 20 rounds of H-BLUP selection. Results showed that bias in VC estimates using H-matrix increased when the proportion of genotyping best individuals increased from 10% to 30%, and EBVs of genotyped animals were also biased upwards in H-BLUP. In contrast, when genotyping was random, much less bias in VC estimates and EBVs were observed. Using VC estimated from traditional animal models in genetic evaluation by H-BLUP reduced the bias in EBVs, but did not remove it completely. Using base population allele frequencies to construct G did not improve VC estimation. We conclude that the bias in VC estimates from using H matrix is due to selective genotyping, and using base population allele frequency would not help remove the bias.

57

Genome-wide association study for reproduction traits in Holstein cattle

Ms. Aline R Guarini[1], Dr. Mehdi Sargolzaei[2], Dr. Valdecy Rocha da Cruz[1], Dr. Angela Cánovas[1], Dr. Filippo Miglior[1], Dr. Flavio Schenkel[1]
[1]CGIL - Department of Animal Biosciences, University of Guelph , [2]The Semex Alliance

With the evolution of mapping studies, association analysis using linkage disequilibrium information within the whole population is now performed to help identify relevant genomic regions associated with traits of economic importance. A genome-wide association study (GWAS) was carried out using a single SNP mixed linear model, including a polygenic term fitting the SNP based genomic relationship matrix, for reproduction traits in Holstein cattle. Estimated breeding values and genotypes from 9,262 bulls genotyped with Illumina Bovine SNP50 BeadChip were obtained from the Canadian Dairy Network (Guelph, ON). Only the 29 bovine autosomes were considered in the analyses. Sixteen traits, including female fertility and calving traits (i.e. age at first service, cow's days open, and heifer's and cow's, 56-day non-return rate, days from first service to conception, days from calving to first service, sire calving ease, calving ease, calf survival, and sire calf survival), were analyzed for association with SNPs. Genomic regions of interest were determined based on the presence of multiple significant neighboring SNPs identified in a chromosomal region. A total of 10 genomic regions on seven chromosomes were significantly detected. Strong evidence for presence of QTL affecting reproduction traits was observed on chromosome 18. The significant SNPs within each region were submitted to NGS-SNP and mapped to the nearby genes for subsequent functional analysis using Ensembl Comparative Genomics Resources. Based on the significant SNP-trait associations and genomic regions determined, along with Gene Ontology (GO) analysis, several genes and microRNAs with relevant functions regulating the traits under study were identified (e.g. NHEJ1, FGF6, CCND2, NF2 , TSKS, ZMYND15, FLT3LG, bta-mir-150 and bta-mir-2900).

58

Diallel analysis of top size in carrot (Daucus carota, L.) using biplots and image-based phenotyping

Sarah D. Turner[1], Nathan D. Miller[2], Edgar P. Spalding[2], Philipp W. Simon[3]
[1]Dept. of Horticulture, University of Wisconsin-Madison, [2]Dept. of Botany, University of Wisconsin-Madison, [3]Dept. of Horticulture & USDA-ARS Vegetable Crops Research Unit, University of Wisconsin-Madison

Carrot crop establishment is limited by erratic germination, poor seedling growth, and delayed canopy closure, resulting in poor competition with weeds and high management costs. Varieties with rapid top growth are one option to improve weed management, but little is known about the inheritance and genetic basis of top size in carrot. This project aims to determine the genetic components contributing to top size using a diallel mating design and to develop an image-based phenotyping platform for carrot morphology, which will facilitate future breeding efforts and reduce the time required for data collection. Six diverse carrot inbred lines were crossed, including reciprocals, in Madison, WI in 2014 and 2015. F1 progenies and parents were grown out in a randomized complete block design (RCBD) with two blocks in El Centro, CA (2014, 2015) and in Hancock, WI (2015). Midseason and harvest measurements were taken for canopy height, canopy width, shoot biomass, and root biomass. Data was analyzed using Griffing's Method I, Model I to estimate general combining ability (GCA), specific combining ability (SCA), and reciprocal effects. Results were represented graphically using the GGE biplot method, which allows visualization of heterotic groups, superior hybrid combinations, and potential gene action. Preliminary results suggest that both non-additive and additive gene action influence height, width, and biomass, with non-additive effects playing a larger role. Bayesian methods are currently being explored as an alternative to traditional means of analysis. Conclusions from this research will provide valuable insight into the inheritance of top size traits and inform selection strategies.

60

Trait-specific long-term consequences of genomic selection in beef cattle

Dr. Haroldo H.R. Neves[2], Dr. Roberto Carvalheiro[1], Dr. Sandra Queiroz[1]
[1]*Departamento de Zootecnia - FCAV - Unesp,* [2]*Gensys Consultores Associados Ltda.*

Simulation studies can be useful to address long-term consequences of selection schemes as applied to real breeding programs, helping to identify effective strategies to enable genetic gain and maintain genetic diversity in longer time horizons. The aim of this study was to evaluate the impact of application of genomic selection (GS) in beef cattle, by simulating scenarios mimicking situations of particular importance to these populations. Forward-in-time simulation was employed to mimic a population with pattern of linkage disequilibrium (LD) close to that verified in real beef cattle populations. Different scenarios of genomic selection and traditional BLUP selection were simulated for 15 generations, mimicking selection of female reproduction, meat quality and growth traits. For GS scenarios, an alternative selection criterion was simulated (wGBLUP), intended to enhance long-term gain by attributing more weight to low-frequency favorable alleles. Marker-based estimators of inbreeding were contrasted to conventional pedigree-based estimators. GS allowed genetic progress up to 40% greater than BLUP under selection for female reproduction and meat quality traits, while no benefit was verified under selection for a beef cattle growth trait. The alternative criterion wGBLUP did not increased long-term response, while it was beneficial to reduce inbreeding rates and loss of favorable alleles. Estimation of inbreeding using marker information based on the concept of runs of homozygosity outperformed the remainder strategies to estimate homozygosis due to identity-by-descent. Large benefits were envisaged for GS over traditional selection for scenarios mimicking selection for meat quality and female reproduction. There was evidence that larger advantage can be expected for GS compared to BLUP when the selected trait is under less polygenic background and that attributing more weight to favorable alleles of low-frequency can contribute to reduce inbreeding rates and loss of favorable alleles in GS.

61

Integrations of risk factors, copy number variants, and gene expression to predict survival of breast cancer patients

Agustin Gonzalez-Reymundez[1], Gustavo de los Campos, Ana Inés Vázquez
[1]*Michigan State University*

Breast cancer (BC) is the second most common type of cancer and a major cause of death in women. Commonly, BC patients are assigned to risks groups based on the combination of prognostic and prediction factors (e.g. patient age, tumor size, tumor grade, hormone receptor status, etc.). While this approach is able to identify risk groups with different prognosis, even within such groups, patients are highly heterogeneous in their response to treatment. In this study, we assess whether the accuracy of prediction of BC survival can be improved, by considering whole-omic profiles for gene expression (GE) and copy number variants (CNV). We describe a modeling framework that is able to incorporate clinical risk factors, high-dimensional omics profiles and interactions between omics and non-omic factors (e.g., treatment). We used the proposed modeling framework and data from METABRIC to assess the impact on the accuracy of BC patient survival predictions when omics and omic-by-treatment interactions are taken into account. Our analysis shows that omics and omic-by-treatment interactions explain a sizable fraction of the variance of survival time that is not explained by commonly used clinical covariates. Furthermore, adding omics and omic-by-treatment interactions led to a relevant increase in prediction accuracy in cross-validations. The sizable interaction effects observed, together with the increase in prediction accuracy, suggest that whole-omic profiles could be used to improve prognosis prediction among BC patients.

62

Comparison between haplotype-based and single SNP-based methods for genomic predictions in Holstein cattle

Mrs. Seyedehzahra Karimi[1], Dr Mehdi Sargolzaei[2], Professor Flavio Schenkel[1]
[1]*Centre for Genetic Improvement of Livestock, University of Guelph,* [2]*Semex Alliance*

Commercial chips mostly lack SNP with very low minor allele frequency due to the ascertainment bias in their design. Therefore, single SNP-based genomic relationship matrix (Gsnp) is expected to put more emphasis on relationships due to distant ancestors. Additionally, Gsnp has less ability to distinguish identity by state from identity by descent alleles compared to a multilocus haplotype-based relationship matrix (GH) that can better detect identity by descent alleles and more recent relationships. The aim of this study was to compare the prediction accuracy using Gsnp vs. GH in North American Holstein cattle. Various traits with wide range of heritabilites, including protein percentage (PRO%; h^2=0.370), fat percentage (FAT%; h^2=0.370), conformation (CONF; h^2=261), daughter fertility (DF;h^2=070), cow first service to calving interval (FSTCc; h^2=0.077), cow calf survival (CSc; h^2=0.023), herd life (HL; h^2=098) and Foot Angle (FAN; h^2=0.109), were analyzed. GH was calculated after converting haplotypes to pseudo SNPs. The training set consisted of proven bulls born before 2008. Observations for the training group were de-regressed EBVs (dEBV). The validation set included bulls that were born in 2008 and 2009 with an official proof in 2014. Correlation between genomic estimated breeding values and dEBV2014 was calculated as a proxy of prediction accuracy. The prediction accuracies for PRO%, FAT%, conformation, and FAN were not improved using GH compared to Gsnp. However, the prediction accuracy was improved by 0.7, 3.1, 2.9 and 1.9 points for CSc, HL, DF and FSTCc, respectively. We concluded that genomic selection based on GH instead of Gsnp seems to improve prediction accuracy for traits under recent selection, which were all lowly heritable in this study. Further studies are needed to investigate the effect of recent selection, heritability and genetic architecture of the traits on the performance of haplotype-based models.

63

Prospects for genomic selection in macadamia, an outcrossing perennial rainforest tree

Ms Katie O'Connor[1], Dr Ben Hayes[1], Dr Craig Hardner[1], Dr Mobashwer Alam[1], Dr Andrzej Kilian[2], Dr Bruce Topp[1]

[1]*University of Queensland,* [2]*Diversity Arrays Technology*

Macadamia is a long-lived rainforest tree native to the central east coast of Australia, and is grown commercially for its edible nuts. In current macadamia breeding programs, both progeny and regional variety trials take at least 8 years each to obtain estimates of yield of nuts and tree growth. Thus, the selection time of elite candidates is very long. Genomic selection (GS) is a useful tool for plant breeding programs, particularly perennial trees, to increase the rate of genetic gain and reduce breeding cycles. We therefore aimed to use GS to accelerate genetic improvement in our breeding program. A preliminary study employing GS in macadamia has been conducted using 198 genotypes, 10,435 SilicoDArT and 3,139 SNP markers. BLR and rrBLUP models gave average validation correlations of 0.59 to 0.87 across various yield and tree growth traits. These data are being used to determine linkage disequilibrium decay, effective population size and genomic heritability for a more in-depth investigation of the potential to use GS in macadamia. The measurements will be used to estimate the required number of genotypes to produce the highest accuracies for predicting yield traits such as nut in shell yield and kernel recovery. Additionally, we report on the design of a genome-wide association study using extensive phenotypic measurements of components of yield and tree growth from the progeny field trial. Outcomes of this study will include predictors of yield, genomic predictions for yield traits that can be employed in the next generation of the macadamia breeding program, and evaluation of strategies of implementing GS into the program.

64

The effects of alternative selective regimes on the genetic basis of adaptation to novel condition in experimental populations

Dr. Yuheng Huang[1], Dr. Aneil Agrawal[2]
[1]*University of Wisconsin-Madison,* [2]*University of Toronto*

Understanding what affects the adaptive potential of the populations is a major goal in evolutionary biology. The potential is determined by the genetic variation available for adapting to novel environments (i.e., the number and the level of variation of the loci that respond to novel selection). As environmental heterogeneity helps maintain genetic variation, it can be an important factor sharping the genetic basis for future adaptation, given that the variation maintained by environmental heterogeneity is related to the novel selection. Here we examine the effects of environmental heterogeneity on the genetic basis of adaptation to novel selection using replicated experimental Drosophila melanogaster populations. The populations had previously evolved for ~100 generations under one of four selective regimes: constant salt-enriched larval medium, constant cadmium-enriched larval medium, and two heterogeneous regimes that vary either temporally or spatially between the two media. Replicates of these experimental populations were subjected to a novel heat stress while being maintained in their original larval diet selection regimes. Within ~20 generations of heat selection, there was evidence of adaptation among different diet regimes for female productivity, with heterogeneous populations showing a larger adaptive response than homogeneous populations. There was less evidence of adaptation overall for male siring success, and populations in constant cadmium diet regime shown the largest response for this trait. We resequenced the heat-selected populations and their pre-selected control populations to infer loci responded to heat selection. With the sequence data before and after heat selection for different diet regimes, this case study is a direct test on whether the genetic variation maintained by environmental heterogeneity contributes to the adaptation to novel conditions.

65

Insights into the genetic architecture of bioenergy traits in eastern cottonwood from a genome-wide association study including common and rare genetic variants

Annette M. Fahrenkrog[1], Leandro G. Neves[1], Marcio F.R. Resende Jr.[1], Christopher Dervinis[1], William B. Barbazuk[1], Matias Kirst[1]
[1]*University of Florida*

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 425 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 seven biomass growth and wood composition traits phenotyped in the population. Including rare variants in this GWAS increased the number of associations identified by 57%. These results highlight the importance of including low frequency variants in GWAS of highly diverse tree species for a better understanding of the genetic architecture of complex traits. 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.

66

On curious properties of genomic relationship matrices in mixed models

Prof. Bruce Tier[1], Prof. Karin Meyer[1]
[1]*A.G.B.U, University of New England*

Genomic relationship matrices (G), constructed from allele counts of genomic markers, have become an integral part of mixed model equations in genetic evaluation, replacing or supplementing the classic, pedigree based numerator relationship matrix, A. Various methods to determine G are available, differing in assumptions about allele frequencies, allele coding used and approaches to modify G so that it has full rank and is 'comparable' to A. We examine simple relationships between selected formulations for G which look rather different but result in equivalent predictions of genetic merit. Firstly, predictions are invariant to adding a (small) constant α to all elements of G, i.e. replacing G with G+αJ where J is a matrix with all elements equal to unity. Whilst the pertaining sampling variances (and thus accuracies of evaluation) derived from the inverse of the coefficient matrix differ they are readily adjusted for the constant added. We thus propose to replace the more common practice of adding a constant to the diagonal elements of G to alleviate rank deficiencies with that of adding αJ. Secondly, predictions using G calculated from 'centered' allele counts, e.g. G=(M-2P)(M-2P)'/d (with M the matrix of allele counts and P the corresponding matrix of frequencies), are equivalent to those assuming allele frequencies of 0.5. Subtracting unity so that genotypes are coded as -1, 0 and 1, i.e. calculating G=(M-J)(M-J)'/c , then allows elements of G to be determined simply by counting the number of identical and opposing homozygotes for each pair of individuals. This involves integer and logical operations only which is computationally far less demanding than the substantial number of floating point operations required when using centered allele counts. An approximate adjustment of sampling variances obtained for this alternate form of G is feasible, with its quality depending on variability in allele frequencies.

67

Genetic evidence of mating for height and body mass index in humans

Dr. Matthew Robinson[1], Dr. Aaron Kleinman[2], Dr. Mariaelisa Graff[3], Dr. Anna Vinkhuyzen[1], Dr. David Couper[3], Dr. Michael Miller[8], Dr. Wouter Peyrot[4], Dr. Abdel Abdellaoui[4], Dr. Brendan Zietsch[5], Dr. Ilja Nolte[6], Dr. Jana van Vliet-Ostaptchouk[6], Prof. Harold Snieder[6], The LifeLines Cohort Study[6], Genetic Investigation of ANthropometric Traits (GIANT) consortium, Dr. Sarah Medland[5], Prof. Nicholas Martin[5], Prof. Patrick Magnusson[7], Dr. William Iacono[8], Dr. Matt McGue[8], Prof Kari North[3], Dr Jian Yang[1], Prof Peter Visscher[1]

[1]University of Queensland, [2]23andMe Inc., [3]University of North Carolina, [4]VU University, [5]QIMR Berghofer Medical Research Institute , [6]University of Groningen, [7]Karolinska Institutet, [8]University of Minnesota

In humans, individuals exhibit a preference for pairing with those who share similar quantitative phenotypes, common disease risk, behaviour, and personality. In people, assortative mating can arise from direct assortment based on phenotypic characteristics, partner interaction and convergence in phenotype over time, or because individuals pair on social or environmental background, referred to as social homogamy. Despite the fact that phenotypic similarity between partners for traits such as height was first quantified over a century ago, distinguishing among these mechanisms is a long-standing question and could not be addressed in previous studies because of the many confounding factors that affect partner similarity. Here, we use a new experimental design and exploit the existence of large-scale genetic data to quantify, for the first time we believe, the degree to which phenotypic assortment for height and body mass index creates a correlation at trait-associated loci among partners. We show that assortative mating for height and body mass index occurs through people directly choosing each other based on their phenotypes in two independent samples of 5,000 and 11,000 spousal pairs. Variation in realised mate choice for height and BMI is associated with common loci, with enrichment at trait-associated loci for height reflecting a shared genetic basis of mate choice and the trait. Our results imply that mate choice in humans affects the genomic architecture of complex traits in the population.

69

Prediction of growth curves in pigs using a hierarchical Bayesian model

Dr. Naomi Duijvesteijn[1], Msc. Roos H Vogelzang[1], Dr. Luc Janss[2], Dr. Egbert F Knol[1]
[1]*Topigs Norsvin Research Center* , [2]*Molecular Biology and Genetics, Aarhus University*

The objective of the study was to derive breeding values for the a, b and c parameter of the Gompertz curve using a Bayesian method and validate them using pedigree or genomic information. Longitudinal data on body weight were used from 3,499 Norwegian Duroc pigs. The pigs were also genotyped with the porcine 60K SNP chip. For validation, 1,110 of the youngest animals were used for prediction of age at weight point 40, 60, 80, 100 and 120 kg. The analysis was performed using a hierarchical model in Bayz (program). In the first level, random animal effects for each pig were estimated per growth parameter (a: asymptotic value, b: point of inflection, and c: decay parameter). The second level models fixed and random effects like sex, litter and polygenic or SNP effects on the animal a, b, and c effects. A total of 150.000 cycles with a burn-in of 50.000 was used to determine the breeding values. Heritability estimates from the posterior means for the a, b and c parameter were 0.66 (0.026), 0.32 (0.027) and 0.81 (0.034) using a Bayesian Variable Selection (BVS) model. Two models were evaluated for their predictive abilities: BLUP using pedigree and BVS using SNPs. Accuracies for BLUP were 0.08, 0.11, 0.13, 0.16, 0.19 for age at weight point 40, 60, 80, 100 and 120 kg. Accuracies for BVS model were higher for all weights; 0.18, 0.25, 0.29, 0.33, 0.36 respectively. Precision breeding for the shape of growth curves using the described method, can improve the efficiency of pork production, and a genomic approach has significant advantage compared to pedigree-BLUP.

70

Predicting direct and indirect genetic effects on survival time in laying hens using repeated observations

Mrs. Tessa Brinker[1], dr. Esther D Ellen[1], dr. Roel F. Veerkamp[1], dr. Piter Bijma[1]
[1]*Wageningen University and Research – Animal Breeding and Genomics Centre*

Minimizing bird losses is important in the poultry industry. Selection against mortality is challenging because heritability is low, censoring is high, and survival depends on social interactions. With cannibalism, mortality depends on Direct Genetic Effects (DGE), i.e. effects of an individual's genotype on its own phenotype and Indirect Genetic Effects (IGE), i.e. effects of an individual's genes on trait values of its social partners. Studies using DGE-IGE models focussed on analysing survival time where censored records were considered as exact and it was assumed that IGE were continuously expressed, even after death. However, dead animals no longer express IGE, and IGE should therefore be time-dependent in the model. Neglecting censoring and timing of IGE expression may reduce accuracy of estimated breeding values (EBV). Thus, our aim was to improve EBV accuracy for survival time in layers showing cannibalism.

We considered four DGE-IGE models to predict survival time: one model analysing survival time and three models analysing survival as a repeated binomial trait. It was also tested whether EBV were improved by including timing of IGE expression in the analyses. Approximate EBV accuracies were calculated by cross-validation. The data contained 13 repeated survival records of two White Leghorn layer lines W1 and WB.

Compared to analysing survival time, EBV accuracy was reduced when timing of IGE expression was included in the model. EBV accuracy was higher when repeated measures were used. Using repeated measures instead increased EBV accuracy by 10 to 21% for W1 and 2 to 12% for WB.

Our results suggest that EBV accuracy for survival time in laying hens can be improved using repeated measures models. This is important because more accurate EBV contribute to increased selection response.

71

Bias of estimated allele substitution effects due to dominance, for two models of gene-action

Mr. Pascal Duenk[1], Dr.ir. Mario P.L. Calus[1], Dr.ir. Yvonne C.J. Wientjes[1], Dr.ir Piter Bijma[1]
[1]*Wageningen University*

Dominance is common for traits important in agriculture. Most genomic prediction models, however, specify only average effects of alleles, because the average effect of an allele includes the heritable component of the dominance effect. Nevertheless, according to literature, models that explicitly account for dominance yield more accurate estimates of breeding values. The mechanism underlying this observation is unknown. The objective of this study was to investigate bias in estimates of allele substitution effects (α) in the presence of dominance, using either an additive (A) model or an additive plus dominance (A+D) model. We considered a large population in Hardy-Weinberg Equilibrium (HWE), from which a sample of N individuals was taken for analysis. For different samples sizes N and allele frequencies p, the expectation of α-hat was computed by multiplying the probabilities of all possible genotype distributions with their least squares estimator of α, and summing these products. The bias is the difference between E[α-hat] and α. Results show that α-hat from the A model was much more biased than α-hat from the A+D model. Bias was larger in smaller samples and when p was more extreme. The bias originated from sampling deviations of genotype frequencies from the exact HWE proportions and to lesser extent from sampling deviations of the population allele frequency. The AD model showed less bias and lower MSE because the effect of sampling deviations of genotype frequencies on α-hat was reduced. The A+D model only showed bias when one of the homozygous genotype classes was not sampled. In conclusion, small deviations from exact HWE in the sample can cause biased estimates for α in the presence of dominance. Our results may explain the greater accuracy of genomic prediction models that explicitly model dominance, particularly when reference populations are small and the distribution of QTL-allele frequencies is U-shaped.

72

Which individuals to phenotype? Optimal design of reference population for genomic selection while maintaining genetic diversity

Sonia E Eynard[1,2,3], Denis Laloe[1], Pascal Croiseau[1], Mario PL Calus[2], Sebastien Fritz[1,4], Gwendal Restoux[1]
[1]GABI, INRA, AgroParisTech, Universite Paris-Saclay, [2]ABGC, Wageningen UR, [3]CGN, Wageningen UR, [4]Allice

With the development of genetic tools and Genomic Selection (GS) an increasing number of individuals are systematically genotyped with commercial SNP chips, giving access to more information for selection. In this genomic era, when doing selection for genetic improvement and/or conservation, the cost-limiting factor becomes phenotyping. In fact, knowing all the genotypes the question becomes which individuals should be phenotyped to update the reference population for GS when trying to insure both genetic progress and diversity conservation. Methods, like Optimal Contribution (OC) selection, have been developed for such purposes and allow selecting groups of individuals maximising genetic merit while minimising average relationships. Our objective was to select individuals to optimise the reference population design for GS. We had 14052 genotypes, Illumina BovineSNP50 BeadChip ® or imputed, and phenotypes for milk production of the French Montbéliarde cattle population. Using GS we predicted performances of a group A2, potential individuals to join the reference population, based on information from A1, initial reference population. With these predictions, we selected groups of A2 individuals (100, 200, 500, 1000 or 2000) to update the reference population (A1+2). This selection decision was made i) randomly, ii) using truncation selection or iii) OC. Then A1+2 was used to predict the performances of V, validation population. Prediction accuracy, i.e. correlation between observed and estimated performances for V, and genetic diversity of A1+2 from estimators based on allele frequencies were measured for each scenario. Even though the differences in genetic diversity were small across the scenarios, selection based on OC consistently resulted in higher amount of diversity while maintaining genetic merit of A1+2 and prediction accuracy for V (on average 0.33) to similar values between optimised scenarios and truncation. In conclusion, OC shows some potential to optimally design reference populations to simultaneously maximise genetic merit and diversity in GS.

73

Candidate gene identification for maize gibberella ear rot disease resistance using genomic and transcriptomic approaches

Dr. Aida Z Kebede[1], Anne Johnston[1], Danielle Schneiderman[1], Whynn Bosnich[1], Dr. Lana M Reid[1], Dr. Linda J Harris[1]
[1]*Ottawa Research and Development Center, Agriculture and Agri-Food Canada*

To improve breeding efficiency for Gibberella ear rot (GER) resistance and gain a deeper understanding of the host-pathogen interaction at the molecular level, advances in next generation sequencing technologies such as Genotyping-by-Sequencing (GBS) and RNA-seq offer powerful tools. Using GBS and a bi-parental recombinant inbred line population, we mapped GER disease resistance quantitative trait loci (QTLs) and significantly improved the accuracy and precision of identified chromosomal regions. Notably, the QTL regions were flanked by a much narrower marker-to-marker distance which could in turn significantly reduce unwanted linkage drag when performing marker-assisted selections. To further characterize the trait, the transcriptional changes observed in maize kernel tissues during the early stages of GER infection were studied in the same inbred lines using RNA-seq. By comparing QTL regions with gene expression data, we identified genes involved in modulation of the plant hormone jasmonic acid, signaling for cell wall modification, cell detoxification, biosynthesis of pathogenesis related proteins and phytoalexin as candidate resistance genes. We used droplet digital PCR to validate gene expression profiles in a broader range of treatments. To conclude, whole genome gene expression profiling using RNA-seq is an invaluable tool to understand and characterize disease resistance through identification of genes which have the potential to serve in marker assisted selection and/or genome editing and rapid deployment of resistant genotypes, especially when gene expression profiling information is coupled with QTL mapping.

74

Within-family genomic selection in aquaculture breeding programmes

Prof. Miguel Toro[1], Dr Maria Saura[2], Dr Jesus Fernandez[2], Dr Beatriz Villanueva[2]
[1]*Universidad Politécnica de Madrid,* [2]*INIA, Madrid*

In classical aquaculture breeding programmes selection within families cannot be applied for traits that cannot be recorded on the candidates. However, this problem can be overcome if genomic information is available. In this study we have investigated within-family genomic selection under two scenarios: 1) base populations in global linkage equilibrium; and 2) base populations with the maximum possible linkage disequilibrium. Populations composed of 50 families were simulated. The genome consisted of 10 chromosomes of 1 Morgan each, with all combinations of 5, 10, 20, 50 and 100 QTL and 5, 10, 20, 50 and 100 markers per chromosome. Two heritabilities (h^2 = 0.2 and 0.4) were considered. Within each family 100 full-sibs were genotyped and phenotyped (training set) and 100 additional sibs were only genotyped (testing set). For h^2 = 0.4, in scenario 1 the correlation between true and genomic values ranked from 0.57 (5 QTL/5 markers) to 0.64 (100 QTL/100 markers) for the training set and from 0.43 to 0.55 for the testing set. For h^2 = 0.2, corresponding figures were 0.42 and 0.48 for the training set and 0.33 and 0.41 for the testing set. The selection differential applied in the testing set was similar to that that would have been applied if the trait could have been recorded and phenotypic within-family selection could have been practiced. In scenario 2, the correlation between true and genomic values was as high as 0.97 (training set) and 0.95 (testing set) even for the case of having only 5 QTL and 5 markers per chromosome. This implies that the selection differential applied in the testing set would be up to 30% higher than the phenotypic selection differential that would have been applied if the trait could have been recorded. Intermediate situations of linkage disequilibrium were also investigated.
ACKNOWLEDGEMENTS**:** The research leading to these results has received funding from the European Union's Seventh Framework Programme (KBBE.2013.1.2-10) under grant agreement n° 613611.

75

Factors affecting the estimation of the genetic correlation between populations

Ms. Yvonne C.J. Wientjes[1], Piter Bijma[1], Pascal Duenk[1], Mario P.L. Calus[1]
[1]*Animal Breeding and Genomics Centre, Wageningen University and Research*

The genetic correlation between populations determines the benefit of using information across populations in genomic prediction. The true genetic correlation is the correlation between allele substitution effects of quantitative trait loci (QTL) underlying the trait, which might differ across populations. In practice, the QTL and their effects are unknown. Therefore, the genetic correlation is generally estimated using information of single-nucleotide polymorphisms (SNPs) that are expected to be in linkage disequilibrium (LD) with the QTL. This LD might differ across populations, which may affect the estimated genetic correlation between populations. Moreover, the variance explained by a QTL can differ across populations, due to differences in allele frequencies and effects. The aim of this study is to investigate the effects of differences in LD and allele frequencies across populations on the estimated genetic correlation between populations. Two populations separated for several generations were simulated in the QMSim software. Allele frequencies of SNPs and QTL were either similar or different in both populations. Phenotypes were simulated by selecting QTL allele frequencies following a U-shape distribution, and by sampling effects of QTL from a multi-variate normal distribution using genetic correlations ranging from -1 to 1 between the populations. Allele frequencies of selected SNP markers followed a uniform distribution. Since the populations separated several generations ago, differences in LD pattern were present. A high or low consistency in LD between QTL and SNPs across populations was obtained by using either a high or low SNP density. The genetic correlation was estimated in a multi-trait GBLUP model for the different scenarios. Results will show the effect of differences in allele frequency and LD patterns between populations on the estimated genetic correlation.

76

Consequences of model misspecification on REML estimates of heritability using genomic models

Dr. Beatriz Cuyabano[1], Dr. Peter Sørensen[1], Dr. Anders Christian Sørensen[1], Dr. Gustavo de los Campos[2], Prof. Dr. Daniel Sorensen[1]
[1]Aarhus University, [2]Michigan State University

Genomic models that incorporate dense marker information are used for prediction and inferences. In the latter case, when the genomic model does not represent correctly the variance covariance structure of the data, the likelihood is misspecified and this can give rise to incorrect inferences. Using simulations, we studied the sampling properties of restricted maximum likelihood (REML) estimates of heritability for different scenarios, each replicated 100 times, in a population of 2,000 individuals. The scenarios comprised 20k single nucleotide polymorphisms (SNPs), of which 100 were assigned as quantitative trait loci (QTL), using unrelated or related individuals, and with different genetic architectures regarding the minor allele frequencies (MAF) of the markers and of the QTL. When individuals were related, average REML estimates did not show detectable bias, regardless of the degree of linkage disequilibrium (LD) between QTL and markers. However when individuals were nominally unrelated, results were highly dependent on the degree of similarity between the genotypic distributions of markers and QTL. When the genotypic distributions of QTL and markers were the same (allele frequencies either all common with MAFs between 0.05 and 0.5, or all rare with MAFs between 0.01 and 0.03), and the QTL were included among the set of markers, average REML estimates did not show sign of bias. This was not the case when the genotypic distributions of QTL and markers differed. In this case a downwards bias was observed. We developed some theory that shows how the addition of marker information affects the likelihood function and that helps to interpret the simulation results. The study indicates that inferences using genomic models must be interpreted with caution.

77

A double repeatability model to evaluate shell quality in layer chicken

Dr. Anna Wolc[1,2], Dr Jesus Arango[1], Dr Petek Settar[1], Dr Neil P. O'Sullivan[1], Dr Jack C.M. Dekkers[2]
[1]*Hy-Line International,* [2]*Iowa State University*

Shell quality is one of the most important traits in layers. Proper consideration of multiple records increases accuracy of breeding values thus genetic improvement of shell quality. A total 81,646 dynamic stiffness records of 24,679 brown egg layers from 4 generations were analyzed herein. Across generations, data were collected up to four ages (26-86 weeks) with repeated records at each age. Five models were compared: animal model with permanent environmental (pe), model with age specific pe, model with both overall and age specific pe, model with age specific pe and age specific residual variances, multi-trait model with each age group as a separate trait. All models included the fixed effects of hatch, and of testing station and age of hen within age group. Analyses were performed in ASReml3. The first three models resulted in similar estimates of heritability of 0.36, 0.32, and 0.31, but repeatability was higher for the model with age specific pe (0.59) than for model with overall pe (0.47). When both pe effects were fitted, the effect within age captured most of the pe variation. Allowing for heterogenous residual variance further improved the fit of the model. The estimates of heritability and repeatability from multi-trait model were 0.43 and 0.61 at age1; 0.44 and 0.61 at age2; 0.31 and 0.50 at age3; 0.21 and 0.53 at age4. The estimates of genetic correlations ranged from 0.27 between ages 1 and 4 to 0.87 between neighboring ages whereas the estimates of pe correlations were between 0.29 and 0.76. The model fitting pe within age and heterogeneous residuals between ages resulted in improved fit compared to the model that fits a single pe effect and homogenous residuals thus it is more appropriate for selection than the traditional model, but its fit was inferior to the multi-trait model.

78

Phenotypic effect of recessive embryonic lethal haplotypes on non-return rate in cattle

Dr. Xiaoping Wu[1], Dr. Goutam Sahana[1], Dr. Bernt Guldbrandtsen[1], Dr. Mogens S Lund[1]
[1]*QGG, Aarhus University*

Widespread use of a limited number of elite sires in dairy cattle breeding increases the risk of deleterious allelic variants being autozygous. In a previous study, using genomic data and insemination records, we had identified 3, 7, 7, and 9 haplotypes which led to extra reproductive failure for Danish Holstein, Danish Jersey, Danish Red, and Finnish Ayrshire cattle, respectively. The aim of this study was to characterize the effect of identified recessive embryonic lethal haplotypes on non-return rates. Non-return rates (0 in case of success and 1 in case of failure) were recorded on 35, 56, 100 and 150 days. A total of 2,402,532, 728,937, 473,608 and 1,316,455 insemination records for Danish Holstein, Danish Jersey, Danish Red, and Finnish Ayrshire cattle respectively, were analyzed. Two linear mixed models, the full model and a reduced model, were used for association analysis. In the full model, the parity and the carrier status of the bull and maternal grandsire were included as fixed effect, and the year and month of insemination and maternal grandsire were included as random effect. The reduce model was similar to the full model but the carrier status of the bull and maternal grandsire excluded. Comparison between two models was done by using likelihood ratio test. In addition, a haplotype based association which treats the parity effect as fixed, candidate haplotypes, polygenic effect and residual effect as random is used to estimate the variance explained by these haplotypes.
Results found that for Danish Holstein, Danish Red and Finnish Red Dairy Cattle, the differences of residual variance between two models were significant. It suggests these haplotypes harbor lethal mutations for prenatal death in cattle. However, for non-return rate 100 and 150 days of Danish Jersey, the differences between two models were not significant.

79

A whole genome association study of yearling weight, carcass traits and primal-cuts of Hanwoo (Korean brown cattle)

Dr. Tae Jeong Choi[1], Dr. Mahboob Alam[1], Prof. Sang-Hyon Oh[2], Dr. Jae Gu Lee[1], Mina Park[1], Dr. Byoungho Park[1], Dr. Sidong Kim[1], Dr. Seung Hee Roh[3], Prof. Seung Hwan Lee[4]
[1]National Institute of Animal Science, [2]North Carolina Agricultural and Technical State University, [3]Hanwoo Imrovement Main Center, [4]Chungnam National University

The purpose of the study was to identify genomic regions affecting growth and meat production traits of Hanwoo cattle. A total of 1,374 males were recorded for yearling weight (YWT), carcass weight (CWT), chuck(CHK), shoulder (SLD), brisket and flank (BAF), ribs (RIB), tenderloin (TLN), striploin (STLN), sirloin (SLN), top round (TRND), round (RND), and fore and hind shins (FHS). Animals were genotyped using the Illumina Bovine 50K SNP panel. A genotyping pruning retained 34,753 SNPs for association analysis. The GCTA and MTG2 software packages were used for single marker association testing and heritability (h^2) estimation of the traits, respectively. Genome-wide suggestive associations were decided after Bonferroni correction (P<1.4377E-06). A total of 46 SNPs were detected on BTA14 and BTA6 for YWT, CWT, SLD, RIB, RND, TRND, and FHS; signifying about 29 candidate genes on BTA14 itself. Above traits were also significant for markers closer to two genes i.e., ASPH and RAB2A. Other significant SNPs for CWT and YWT were close to the FAM10B, RB1CC1, and many other genes. Few other genes also deemed closer to the detected SNP for RIB, SLD, FHS, RND, and TRND. However, no QTL was found for chuck, brisket and flank, and loin-cuts. This might have indicated the influences of many undetected QTL with relatively smaller effects for those traits in Hanwoo. The h^2 estimates for YWT and CWT were 0.40 and 0.56, respectively. Most primal cuts were moderate to highly heritable as well, such as 0.21 (BAF) to 0.69 (RND). Among the other traits, the heritability for SLD, STLN, and TRND were 0.63, 0.62, and 0.59, respectively. Given the scarcity in genome-wide association studies for beef primal-cuts, this study essentially suggested a good starting point for selection based on significant markers in the future.
Keywords: Yearling weight, carcass weight, primal-cut, genome-wide association, Hanwoo.

80

Chemotherapy-induced neutropenia is modulated by drug-specific genes

Daniel Gatti[1], Sussane Weber[2], Neal Goodwin[1], Frank Lammert[2], Gary Churchill[1]
[1]*The Jackson Laboratory,* [2]*Saarland University Medical Center*

Neutropenia, a decrease in circulating neutrophil counts, is a common, life-threatening side effect of many chemotherapy regimens. Decreased neutrophil counts are associated with increased risk of infection, increased hospitalization and delays in treatment. While clinical factors such as sex, age and comorbidities are known to influence chemotherapy-induced neutropenia, genetic factors may also influence patient responses. We used a mouse model to demonstrate that genetic factors may play an important role in determining individual patient response. In order to find genes responsible for chemotherapy-induced neutropenia, we dosed Diversity Outbred mice with one of three chemotherapy drugs; doxorubicin, cyclophosphamide or docetaxel and observed a high level of genetically determined variation in neutropenic response. In order to identify genes responsible for neutropenia, we mapped the drug-induced decrease in neutrophil counts after dosing. For doxorubicin, we identified a region on chromosome 5 that contains two ATP-binding cassette genes (Abcb1a, Abcb1b), which are efflux transporters associated with doxorubicin resistance. This association remained even when the mice were treated with granulocyte colony stimulating factor. For cyclophosphamide, we identified a region on chromosome 11 that contains transformation related protein 53 (Trp53), a DNA repair gene and a plausible candidate gene for a DNA alkylating agent. For docetaxel, we identified a gene dense region on chromosome 4. These results suggest that susceptibility to chemotherapy-induced neutropenia may be influenced by different genes for different chemotherapy drugs.

81

Multiple copies of alkane metabolic genes in bacteria

Dr. Hisako Masuda[1], Dr. Peter Tupa[1]
[1]*Indiana University Kokomo*

Alkane monooxygenase (AlkB) catalyze the initial, rate limiting step of alkane metabolism in bacteria. Survey of all published genomes revealed that 65 % of Actinobacteria and 20% of Proteobacteria carry alkB gene on their genomes. Intriguingly, significant numbers of strains carry more than 1 copy of alkB. The goal of this study was to analyze the sequence identities between the alkBs within a single organism and to investigate how strains might have acquired multiple copies of this gene. While we identified few instances of gene duplication and inter-phyla horizontal gene transfer (HGT) events, the HGTs within the phylum seem to play significant roles in acquisition and maintenance of multiple copies.

82

Data structure requirements for estimation genotype-environment interactions when genomic information is available

Dr. Freddy Fikse[1], Dr. Lucy Crooks[2], Prof. Erling Strandberg[1]
[1]SLU, [2]Sheffield diagnostic genetics service

Experiences from international genetic evaluations of dairy cattle suggest the importance of parents having phenotyped progeny in multiple environments to estimate the degree of genotype-environment interaction using a character-state model and when pedigree is used to describe relationships between individuals. High-throughput genotyping enables assessing realized relationships between individuals (based on genomic data) rather than expected relationships (based on pedigree). The aim of this study was to test the hypothesis that use of genomic data reduces the importance of phenotyped progeny in multiple environments. Data were de-regressed breeding values, which are unregressed measures of daughter performance, for three biological traits (protein yield, stature and somatic cells) for Brown Swiss bulls from four populations (environments). In addition, after quality control information on (imputed) genotypes for 45 thousand SNPs was available for all (5240) bulls. A character-state model was fitted, regarding performance in each population (environment) as a different trait, and genetic correlations were estimated by REML for two different data structures and employing two different relationship matrices. Using all de-regressed breeding values, genetic correlations were estimated with either a pedigree-based relationship matrix or a genomic relationship matrix. An alternative data structure was created by removing phenotypic links: for bulls with daughters in multiple environments (populations) we only kept information in the environment the bull had most daughters, and deleted daughter information from the other environments. Again, genetic correlations were estimated using both relationship matrices. Removing phenotypic links resulted in an average decrease of genetic correlation of 0.04, 0.04 and 0.07 for protein, stature and somatic cells, respectively, when pedigree relationships were used. The decrease was smaller (0.00, 0.02 and 0.04 for protein, stature and somatic cells, respectively) when genomic relationships were used. The use of genomic information reduces the demands on data structure to estimate genotype-environment interaction, yet phenotypic links remain important.

83

Including crossbreds in the genomic relationship matrix through utilization of both linkage analysis and linkage disequilibrium

Ms Maja W. Iversen[1], Dr. Eli Gjerlaug-Enger[1], Dr. Eli Grindflek[1], Dr. Marcos S. Lopes[3], Dr. Theo H.E. Meuwissen[2]

[1]Topigs Norsvin, [2]Institute of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, [3]Topigs Norsvin Research Center

In the pig industry, the main aim is to produce good crossbreds for meat production. It has been suggested that when the correlation between crossbred and purebred performance is below 0.7, information from crossbreds may be beneficial when estimating breeding values for the purebred parents of crossbreds. Therefore, it is necessary to explore options to include crossbreds in the (genomic) relationship matrix (G-matrix). The aim of this study is to build a G-matrix that, in an optimal way, can include crossbreds, animals of different breeds, as well as ungenotyped animals. To achieve this, a GLDLA matrix was built, utilizing both linkage disequilibrium (LD) and linkage analysis (LA) information. This is done by first building a gametic relationship matrix, based on genotype probabilities, and then applying adjustments to this gametic relationship matrix to adjust it to a genomic relationship matrix to use in breeding value estimation. The LD part accounts for association between founder haplotype fragments, and the LA part uses markers to directly estimate relationships between animals. This is expected to be beneficial when including crossbreds in the G-matrix because breed relationships are accounted for. To account for different allele frequencies in the different breeds, the founder animals of each breed enter into different genetic groups in the linkage analysis, i.e. founder allele frequencies are estimated within each breed. This also estimates different allele frequencies in the crossbreds because they are linked to the pure lines through the pedigree. Currently, analysis is being performed on real data from 1377 Topigs Norsvin F1 crossbreds, as well as 3238 Dutch Landrace and 3737 Large White purebreds from which the crossbreds originate.

84

When Bayesians encounter Amdahl's law - A statistical perspective of parallel computing for genomic prediction

Dr Xiao-Lin Wu[1], Dr Stewart Bauck[1]
[1]*Geneseek, A Neogen Company*

We are facing challenges brought by unprecedentedly large genomic data. At its core is the process that transforms genome data into breeding decisions. Use of parallel computing techniques is critical to bypass this computational bottleneck.
Bayesian hierarchical models implemented through Markov chain Monte Carlo (MCMC) sampling are commonly used in genomic selection, but its algorithm is computationally sequential. A naïve parallel MCMC is to generate several independent MCMC chains on different processors and then to combine posterior samples appropriately. Assemble of "multiple samples", however, may not always be ergodic to the target (posterior) distribution. Besides, the overall reduction in runtime from parallelism can be limited if every chain has to spend a significant proportion of its time in burn-in. According to Amdahl's law, the speedup in parallel computing is limited by the portion that can't be parallelized. So far, parallel computing algorithms of single MCMC chains have not been made available. An alternative is cascaded parallel computing (CPS), yet its efficiency depends on the tradeoff between the computing and synchronization or communications among nodes.
While the current parallel computing jobs mostly belong to the category of "task-parallelism", a data-parallel ridge-regression BLUP (DP-RR-BLUP) is described hereby, in which SNP effects are estimated using RR-BLUP in each subset and then weighted into the whole-population estimates. The weights are statistically derived that provides exact re-formation of the whole-dataset estimates of SNP effects. This method is useful in three situations: 1) training genomic prediction equations using a dataset that is too large to be handled as a single input, 2) updating genomic prediction equations when new data become available, but without the need to re-train the whole dataset, and 3) Estimate SNP effects from a K-folder cross-validation. In particular, RR-BLUP can be implemented through MCMC so that matrix inversion in each subset is avoided.

85

Use of electronic health record to predict family relationships for genetic epidemiology research

Mr. Xiayuan Huang[1], Dr. Robert Elston[3], Dr. Guilherme Rosa[1], Mr. John Mayer[2], Dr. Ye Zhan[2], Dr. Terrie Kitchner[2], Dr. Page David[1], Dr. Scott Hebbring[1,2]
[1]University of Wisconsin-Madison, [2]Marshfield Clinic Research Foundation, [3]Case Western Reserve University

Diseases are the result of genetic and/or environmental factors. Understanding the etiology of human disease is necessary to advance precision medicine. Pedigree analysis is a longstanding approach to gain insight into the underlying genetic factors in human health. In the last decade, pedigree analysis has largely given way to large population-based studies on samples of independent persons as a result of high-throughput genotyping methodologies and the ability to collect large numbers of unrelated cases and controls. A reverse paradigm shift could occur if the construction and use of family pedigrees could be combined with classical genetic analyses in a high-throughput manner. The goal of this study is to demonstrate that family pedigrees may be rapidly predicted from an electronic health record (EHR) system with no human intervention. Results demonstrate our ability to identify with high probability 173,368 family pedigrees using basic demographic data available in most EHRs. We further show as proof-of-principle that these large patient pedigrees may be applied to genetic research. With the anticipated widespread application of genomic medicine, in combination with methods capable of predicting family pedigrees linked to extensive and longitudinal phenotypic data, a revolution may occur in how human genetic research is conducted for the advancement of precision medicine.

86

SNP discovery in wheat RIL population using genotyping-by-sequencing and genome-wide QTL mapping for plant height

Mr. Waseem Hussain[1], Dr. Stephen Baenziger[1], Dr. Vikas Belmakar[1], Dr. Mary Guttieri[2], Mrs. Amanda Easterly[1], Mr. Jorge Venegas[1], Dr. Jesse Poland[2]
[1]University of Nebraska, Lincoln, [2]Kansas State University

Genotyping-by-sequencing (GBS) is one of the next generation sequencing method that allow sequencing, discovery and genotyping of thousands of SNPs in cost effective manner and quickly. The SNPs generated through GBS can be used to develop the high density linkage maps for precision QTL mapping in wheat. The present project was undertaken with following objectives to: (i) develop high density linkage map in 204 recombinant inbred lines (RILs) derived from contrasting parents Harry and Wesley, (ii) determine the reliability of the newly constructed map with known genes of chaff color and wax/ glaucousness, (iii) compare the newly constructed map with wheat reference genome and related species, and (iii) identify QTLs, digenetic interactions and QTL x environment interaction for plant height. After stringent filtering, a high-density linkage map was constructed. The newly constructed linkage map has 2923 markers distributed on 37 linkage groups and spanned 5269.34 cM with an average distance of 1.79 cM between adjacent markers. The high accuracy and reliability of this map was illustrated by finding and co-localizing the genes for chaff color and wax/glaucousness to correct and previously mapped genomic regions. For plant height in total 18 QTLs were identified across all locations on linkage groups 2DS, 2BL, 3A, 3B.3, 6A.2, 7AL and 7B.2 and phenotypic variance explained by these QTLs ranged from 4.9 to 16.8 %. Digenetic interactions between QTLs was evident, however, none of the interactions were stable across locations. Six QTLs revealed significant interactions with environment and accounted 1.11 to 2.73% of phenotypic variation. Interestingly a major QTL qph.hw.2DS was found in all the five environments. It explained 7.4 to 16.4% of phenotypic variation and height reducing allele for this QTL was contributed Wesley. This QTL may be responsible for the possible cause of plant height variation in the RIL population.

87

Optimizing size and sex ratio of the genotyped population for genomic selection in broilers using information from both genotyped and non-genotyped individuals

Dr. Setegn Worku Alemu[1], Dr. Robyn Sapp[2], Dr. Lie Wang[1], Dr. John Henshall[2], Dr. Rachel Hawken[2], Dr. (senior scientist) Anders C Sørensen[1], Professor Just Jensen[1]
[1]Department of Molecular Biology and Genetics, Aarhus University, Tjele DK-8830, Denmark , [2]Cobb-Vantress Inc., Siloam Springs, AR, 72761, USA

Genomic selection using GEBV based on dense marker is an attractive design for livestock species. Nowadays, breeding programs using genomic information become common for broilers. In broiler the main trait body weight is measured early in both sexes and the heritability of the body weight is high. This might reduce the benefit of genomic information. However; the benefit of genomic information for broiler can be maximized by including additional traits such as residual feed intake to body weight.
A simulation study was conducted to optimize breeding design using genomic information for body weight and residual feed intake of broilers. The simulation was run for forty rounds of selection using ADAM software. The parents (13 sires out of 2600 candidates and 130 dams out of 2600 candidates) were selected based on EBV from BLUP for the first twenty rounds of selection, and the subsequent twenty rounds using GEBV from single-step GBLUP. In each round of selection all individuals were measured for body weight and the heaviest 20% of individuals were measured for residual feed intake. We tested the following factors: the size of the genotyped population (130, 260, 520, 1040) and the sex ratio for the genotyped population (0% male, 25% male, 50% male, 75%male, 100%male). We found significant interaction between the size of the genotyped population and sex ratio. For smaller training size (130), genotyping males only results in higher response to selection while for larger training size (1040) genotyping both sexes is a better option. The reasons for this interaction may be due to difference in selection intensity for male and female and partial recording of residual feed intake. Thus breeding design for broiler using genomic information should be optimized by the combined effects of the size of the genotyped population and sex ratio of the genotyped population.

88

Modelling the relationship between social interactions and inherited variability

Jovana Marjanovic[1,2], Han A Mulder[1], Piter Bijma[1]
[1]*Animal Breeding and Genomics Centre, Wageningen University and Research,* [2]*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences*

Social interactions are widespread in nature and affect the course of evolutionary processes and response to selection in domestic populations. In QG, we use indirect genetic effects (IGE) to model social interactions. Phenotypic studies, particularly in aquaculture, show that competitive interactions inflate variability of trait values among individuals. Several studies have shown that variability of trait values can have a genetic component, which makes variability a heritable trait. The observed relationship between social interactions and inherited variability hints at an underlying genetic relationship between the two phenomena, of which very little is known. Current IGE models cannot explain effects of competition on variability, while models for inherited variability ignore social interactions. Here we present a QG-model that integrates social interactions and inherited variability. In this model, the focal phenotype, say body weight, is a function of genetic and environmental effect of the focal individual and of the difference in body weight between the focal individual and its social partner, multiplied by a regression coefficient. This regression coefficient, b, depends on a direct genetic effect of the focal individual, representing sensitivity to competition, and an indirect genetic effect of its social partner, representing competitive effect. The b is negative with competition, and positive with cooperation. Simulation results show that our model yields increased variability of body weight for negative b, similar to what is observed in real aquaculture populations. When b becomes positive, variability becomes significantly smaller. The model also shows that social interactions and variability co-evolve, as b can respond to selection. Selection for cooperation will therefore lead to decreased variability. We currently investigate whether ordinary IGE models and models for inherited variability capture the effect produced by our model, and how those models can be extended to capture the genetic effects determining the co-evolution of competition and variability.

89

Tunable combinatorial transcriptional engineering based on CRISPR interference

Katia Tarasava[1], Dr. Rongming Liu[1], Dr. Andrew Garst[1], Dr. Ryan T. Gill[1]
[1]*University of Colorado-Boulder*

Gene expression is precisely regulated in cells to achieve the optimal balance of energy and resource distribution to meet the cell's metabolic needs. When scientists deliberately perturb gene expression for metabolic engineering purposes, it is usually done via a crude approach of massive overexpression or knockout, which often does not produce the desired phenotypes. An improved process of optimization of metabolic flux through a pathway involves making small perturbations in the levels of expression of key enzymes in a combinatorial fashion. We hypothesize that by testing a continuous range of gene expression values (as opposed to only extrema), we can improve pathway flux. We propose a new method for fine-tuning gene expression based on CRISPR interference and activation technologies, where we modulate the level of expression by incorporating mismatches in the gRNA sequence, thus affecting its strength of binding to the target DNA. This method allows rapidly generating and testing many combinatorial expression variants through construction of plasmid-based gRNA arrays. We also develop a mathematical model for prediction of strength of repression/activation based on sequence similarity of the gRNA to the target DNA sequence.

90

Genomic prediction in cattle based on sequence data

Urs Schuler[1], M. Frischknecht[1,2], T.H.E. Meuwissen[3], B. Bapst[1], F.R. Seefried[1], C. Flury[2], D. Garrick[4], H. Signer-Hasler[2], C. Stricker[5], Intergenomics Consortium[6], A. Bieber[7], R. Fries[8], I. Russ[9], J. Sölkner[10], A. Bagnato[11], B. Gredler[1]

[1]Qualitas AG, [2]Bern University of Applied Sciences, School of Agricultural, Forest and Food Sciences HAFL, [3]Norwegian University of Life Science, [4]Iowa State University, [5]agn Genetics, [6]Interbull Center, [7]Research Institute of Organic Agriculture (FiBL), [8]Technische Universität München, [9]Tierzuchtforschung e.V., [10]University of Natural Resources and Life Sciences, [11]University of Milan

In routine genomic predictions in cattle breeding 50k SNP chip data is normally used. However costs for sequencing individuals dropped and imputation to sequence level has been shown to be accurate. Therefore it would be possible to increase the marker density for genomic prediction. This is thought to increase accuracy of genomic breeding values. However to what extent is not clear yet. While simulation studies often promise a relatively high gain in accuracy, in real data often only a minor effect can be observed. We tested different densities of SNP for genomic prediction in a set of about 22,000 Brown Swiss cattle (BSW) which have been imputed to whole-genome sequences (16,184,800 variants). As reference panel for imputation BSW individuals from the 1000 bull genomes project were used. The trait of interest was non-return rate in heifer after 56 days. This trait has a low heritability and as a functional trait in the complex of fertility traits it is of special interest. We performed first a genome-wide association study (GWAS) and identified a clear signal. Within the region of this signal two non-synonymous variants have been associated with the trait.

However we were particularly interested in the different accuracies obtained from different SNP densities. We calculated genomic breeding values based on 2,258 reference individuals and 892 validation individuals. The reference individuals were the same individuals as the ones used to calculate the GWAS. We tested 50k SNP data; top variants from GWAS (50k variants); missense variants (34k variants); LD pruned sequence variants (6M variants) and sequence variants (13M variants). We used a BayesC algorithm. The highest correlation was found using LD pruned sequence data and the lowest using 50k data only. However differences remained relatively small. The same calculations will be done for other traits.

92

Controlling population structure in the genomic prediction of tropical maize inbreds

Danilo H Lyra[1], Ítalo S.C Granato[1], Pedro Patric P Morais[1], Filipe C Alves[1], Anna Rita M Santos[1], Xiaoqing Yu[2], Jianming Yu[2], Roberto Fritsche-Neto[1]
[1]*Department of Genetics, University of São Paulo, Luiz de Queiroz College of Agriculture, Piracicaba, São Paulo, Brazil,* [2]*Department of Agronomy, Iowa State University, Ames, IA, U.S.A.*

Population structure is a confounding parameter that affects genomic predictions, and could lead to biased accuracies. The aim of this study was to compare the prediction accuracy of different correction methods of population structure in the genomic prediction model. Sixty-four inbreds were used in the experimental trial carried out in Anhembi Station in Piracicaba-SP, Brazil, during the second growing season of 2015. The trait evaluated was grain yield. Variance components estimation was performed by REML and the GEBV for individuals was obtained by GBLUP model, using the following four approaches. A model that (a) ignores population structure (GBLUP), (b) includes the admixture coefficients estimated by ADMIXTURE (AD-GBLUP), (c) includes first three PCs (PC-GBLUP), and (d) includes a matrix of zeros and ones based on group clustering calculated in fineSTRUCTURE (FI-GBLUP). The population structure related variables were used as fixed covariates in the model. The data were analyzed with ASReml-R. Prediction accuracy was evaluated using 4-fold cross-validation with 100 replications. The marker matrix was composed by 22K SNPs. The first three PCs accounted for 4.7%, 4.2% and 3.7% of the genetic variance. The inbreds were split into two subpopulations (K). The K_1 consisted of 53 inbreds and K_2 of 11 inbreds. The prediction accuracy varied among the models GBLUP (0.19), PC-GBLUP (0.24), AD-GBLUP (0.20), and FI-GBLUP (0.20). PC-GBLUP model had a relatively good prediction performance. In conclusion, explicitly account for population structure leads to higher accuracies of genomic prediction.

Funding acknowledgement: São Paulo Research Foundation-FAPESP (Process: 2013/24135-2; 2014/26326-2; 2015/14376-8).

93

Patterns of genomic and phenomic diversity in grape and apple

Zoe Migicovsky[1], Jason Sawler[1,2], Kyle Gardner[1], Mallikarjuna K. Aradhya[3], Gan-Yuan Zhong[4], Sean Myles[1]
[1]Faculty of Agriculture, Dalhousie University, [2]Anandia Labs, [3]National Clonal Germplasm Repository, USDA-ARS, [4]Plant Genetic Resources Unit, USDA-ARS

Grape (Vitis vinifera) and apple (Malus domestica) are two of the world's most valuable fruit crops. However, their breeding is limited by their large size, lengthy juvenile phase as well as the expense associated with growing and phenotyping. As a result, these perennial crops are particularly promising for marker-assisted selection (MAS), which would allow plants with desirable genotypes to be selected at the seed or seedling stage. Advances in MAS require genotype-phenotype associations to be established. We performed genome-wide association studies (GWAS) and genomic prediction on over 500 grape and apple accessions from the USDA collection. We find several associations of note including a novel association between the transcription factor, NAC18.1, and harvest date and fruit firmness in apple. We also identify signatures of selection for several traits that overlap with our GWAS results. Given rapid linkage disequilibrium decay in both grape and apple, with sufficient genome-wide coverage there is great promise for high resolution mapping of causal variants in these species which will enable more efficient breeding through MAS.

94

The identification of haplotypes associated with major depressive disorder

David M Howard[1], Lynsey S Hall[1], Jonathan D Hafferty[1], Yanni Zeng[1], Mark J Adams[1], Pippa A Thomson[2], Andrew M McIntosh[1]

[1]Division of Psychiatry, University of Edinburgh, [2]Institute of Genetics and Molecular Medicine, University of Edinburgh

Previous studies of major depressive disorder (MDD) have demonstrated a polygenic and complex genetic architecture, however few variants reach genome-wide significance. Genotype based analyses are unlikely to capture the full variation in the regions surrounding markers, especially those variants which are at lower frequencies and/or those in incomplete LD with typed variants. Partitioning of phased genotype data into haplotype blocks should allow improved tagging of these rarer variants. This study used a mixed linear model to identify haplotypes within these blocks which were associated with MDD. An additional fine mapping procedure was developed, refining the block boundaries, to further elucidate putative causal haplotypes.

The study made use of the Generation Scotland cohort consisting of 21,516 individuals genotyped using the Illumina OMNIExpress BeadChip (706,786 SNPs). Following quality control, phasing was conducting using SHAPEIT v2 (r837) resulting in phased information for 19,994 individuals at 561,125 SNPs. A sliding window approach was taken using 1cM, 0.5cM and 0.25cM windows based on the HapMap phase II recombination rate on GRCh37, with each window sliding along a quarter of the window size respectively. A mixed linear model based association analysis was implemented in GCTA v1.25.0.

The association analysis revealed a genome-wide significant haplotype on chromosome 10 ($P = 8.5 \times 10^{-9}$) with an odds ratio of 2.34. The fine mapping approach revealed a further genome-wide significant haplotype on chromosome 6 ($P = 7.1 \times 10^{-9}$) with an odds ratio of 1.83. The haplotype on chromosome 6 was located within a region previously associated with bipolar disorder. As expected, these rare genome-wide significant haplotypes explained only a small proportion of the total phenotypic variance ($r^2 \approx 2.3 \times 10^{-4}$). However their identification may allow the identification of biologically informative aetiological subtypes for MDD.

95

The epistasis boundary: Linear vs. nonlinear genotype-phenotype relationships

Dr. Serge Sverdlov[1], Prof. Elizabeth A. Thompson[1]
[1]*University of Washington*

Nonlinearity in the genotype-phenotype relationship can take the form of dominance, epistasis, or gene-environment interaction. The importance of nonlinearity, especially epistasis, in real complex traits is controversial. Network models in systems biology are typically highly nonlinear, yet the predictive power of linear quantitative genetic models is rarely improved by addition of nonlinear terms, and association studies detect few strong gene-gene interactions. We find that complex traits satisfying certain conditions can be well represented by a linear genetic model on an appropriate scale despite underlying biological complexity. Recent mathematical results in separability theory determine these conditions, which correspond to three biological criteria (Directional Consistency, Environmental Compensability, and Pathway Redundancy) together making up an Epistatic Boundary between systems suitable and unsuitable for linear modeling.

For genetic traits controlled by limited numbers of loci and alleles, our algorithm enumerates all possible trait structures and finds exact or error-minimizing linearizing transformations by formulating a constrained optimization program. We find that the fraction of possible distinct genetic traits satisfying simple criteria that can be fully or approximately linearized is high for small systems and falls with system complexity.

For nonlinear traits, we introduce a combinatorial classification of types of nonlinearity in a network context. We use this to illustrate how upstream controlling genes, potentially more important to explaining biological function, can be intrinsically harder to detect by GWAS than their downstream controlled counterparts.

96

Additive and non-additive genetic effects for disease resistance in the Pacific White Shrimp Penaeus vannamei

Dr. Héctor Castillo-Juárez[1], Dr. Hugo H Montaldo[2], Dr. Gabriel Ricardo Campos-Montes[1], Ing. Juan Carlos Quintana-Casares[3], Dr. Sonia Soto-Rodríguez[4], M. Sc. Leobardo Montoya-Rodríguez[4], Dr. Miguel Betancourt-Lozano[4], Biol. Alfonso Martínez-Ortega[3], Dr. Rodolfo Lozano-Olvera[4], M. Sc. Alejandra Caballero-Zamora[1]

[1]Universidad Autónoma Metropolitana, [2]Universidad Nacional Autónoma de México, [3]Maricultura del Pacífico SA de CV, [4]CIAD AC

Survival rates from larval stage experimental challenges for two major worldwide diseases in cultured shrimp, Acute Hepatopancreatic Necrosis Disease (AHPND) and White Spot Syndrome Virus (WSSV), were obtained at a large hatchery (Maricultura del Pacífico) in Sinaloa, on the Pacific coast of Mexico during 2015. The objective was to estimate crossbreeding effects, inbreeding depression effect, and heritabilities for disease resistance. A mixed population was formed from purebred and crosses between Ecuadorean shrimp with a history of resistance to WSSV, and shrimp from a Mexican line selected for high growth rate. Additionally, inbred animals were produced from sibling mating. The AHPND trial involved 5,171 animals from 182 sib families (119 sires). The WSSV trial involved 6,231 animals from 181 sib families (118 sires). Results were analyzed using linear mixed models. For AHPND, crossbreeding and inbreeding effects were not significant (P>0.10). For WSSV, Ecuadorean-Mexican lines direct crossbreeding effect was 0.09±0.02 (P<0.01). Heterosis effect was -0.06±0.02 (P<0.01). Crossbreeding maternal and inbreeding effects were not significant (P>0.10). Survival rate direct and maternal heritability estimates for AHPND were 0.12±0.04, and 0.04±0.03, irrespective of whether crossbreeding effects were included in the model. Direct-maternal correlation was negative (-0.86±0.12). Survival rate direct heritability estimate for WSSV was 0.07±0.02 when including the crossbreeding effects in the model, but increased to 0.12±0.02 when these effects were excluded. It is concluded that for AHPND resistance, there are within-line additive genetic differences. For WSSV resistance, there are differences in favor of animals of the Ecuadorian line and additive genetic variation within and between lines. The negative heterosis estimate suggests the presence of a recessive major gene. Genetic parameter estimates for survival time were similar. It is possible to develop lines of shrimp from this population with greater genetic resistance to these diseases using both selection and crossbreeding.

97

The genetic architecture of gene expression in human peripheral blood

Dr. Luke Lloyd-Jones[1], Mr. Alexander Holloway[1], Mr. Andrew Bakshi[1], Dr. Allan McRae[1], Assoc. Prof. Jian Yang[1], Dr. Kerrin Small[3], Dr. Jing Zhao[6], Dr. Andres Metspalu[4], Prof. Manolis Dermitzakis[5], Prof. Greg Gibson[6], Prof. Tim Spector[3], Dr. Grant Montgomery[7], Prof. Tonu Esko[4], Prof. Peter M Visscher[1], Dr. Joseph E Powell[2]

[1]Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, [2]Institute for Molecular Bioscience, University of Queensland, [3]Department of Twin Research and Genetic Epidemiology, King's College London, [4]Estonian Genome Center, University of Tartu, [5]Department of Genetic Medicine and Development, University of Geneva, [6]School of Biology and Center for Integrative Genomics, Georgia Institute of Technology, [7]QIMR Berghofer Medical Research Institute, 300 Herston Road

Understanding how variation in gene expression is altered by changes in genomic structure is fundamental to the study of complex traits and common diseases. We analysed the mRNA levels, measured from blood tissue, of 2,765 individuals to investigate comprehensively the genetic architecture underlying gene expression. Using linear mixed models for expression quantitative trait loci (eQTL) detection, we identified 11,357 cis- and 2,343 independent trans-eQTL at robust study-wide significance thresholds. Using complimentary population- and family-based methods, heritability (h^2) was estimated for 38,624 transcript expression traits. Of these, 13,471 (35%) have an estimated narrow-sense heritability greater than 0.1, with a median across all probes of 0.34. For probes with $h^2>0.1$, eQTLs explained on average 23% of h^2, whereas all genome-wide common (MAF > 0.05) SNPs account for 36% of h^2. Therefore, on average the majority of genetic variance for gene expression is not tagged by common SNPs and the majority of variance that is tagged is due to detected eQTL. Finally, we present evidence that, compared with a meta-analysis, using individual level data results in an increase in power of 13% to detect eQTLs.

98

QTL mapping and genomic selection methods for assisted breeding in the American cranberry

Giovanny Covarrubias-Pazaran[1], Brandon Schlautman[1], Luis Diaz-Garcia[1,2], Lorraine Rodriguez-Bonilla[1], Edward Grygleski[3], Dr. James Polashock[4], Dr. Nicholi Vorsa[5], Dr. Juan Zalapa[1,6]
[1]University of Wisonsin-Madison, [2]Instituto Nacional de Investigaciones Agrícolas, Forestales y Pecuarias, [3]Valley Corporation, [4]USDA-ARS, Genetic Improvement of Fruits and Vegetables Laboratory, [5]Blueberry and Cranberry Research and Extension Center, Rutgers University, [6]USDA-ARS, Vegetable Crops Research Unit

Classical QTL mapping strategies coevolved with genetic markers during the last century. More recently, next generation sequencing technologies have made high-throughput genotyping affordable for minor and understudied crops. Therefore, genome wide association mapping strategies (GWAS) and genomic selection (GS) methods have become a viable option for breeding programs focusing in annual and perennial crops. We the discuss the strengths and weaknesses of classical QTL mapping, association methods, and breeding value predictions using biparental populations in cranberry, which represents a model species for perennial fruit breeding. We used data for fruit yield (FY) and average weight per fruit (AFW) measured across 3 years in three cranberry biparental populations with different degree of relationship. We conducted comparisons among single marker analysis (SMA) and composite interval mapping (CIM) methods based on each population, and GWAS conducted by mixing all populations to compare results among the three mapping methods. Markers providing a signal from classical mapping methods such as SMA and CIM were used to calculate prediction accuracy using multiple linear regression (MLR) and compare against genomic selection using genomic-BLUP (GBLUP) and genomic-bagging-BLUP (bagging-GBLUP). We found the GWAS method to be more powerful to detect higher number of QTLs, but less powerful to detect QTLs from specific populations. Genomic selection provided higher predictive ability than MLR (rGBLUP-FY=0.23 versus rMLR-FY=0.17 and rGBLUP-AFW=0.27 versus rMLR-AFW=0.10), and we found that bagging-GBLUP was able to overcome the excessive shrinkage imposed by the GBLUP methods, providing greater predictive ability than any other method tested in this research for both quantitative traits (rBGBLUP-FY=0.36 versus rBGBLUP-AFW=0.40). The effect of the degree of relationship among individuals and number of markers in the predictive ability for both traits was investigated. The simultaneous use of GWAS and GBLUP will constitute the future of cranberry assisted-breeding strategies potentially applied to other perennial fruit crop species.

99

A novel pedigree-free linkage analysis using multiple regression

Dr. Bujun Mei[1], Dr. Rohan Fernando[2], Dr. Jian Zeng[2], Dr. Hao Cheng[2]
[1]*Hetao College,* [2]*Iowa State University*

Genome-wide association study (GWAS) typically requires a base of linkage disequilibrium (LD) to capture QTL signals. In this study, we tested whether only linkage information can alternatively be based on identifying QTL in the framework of GWAS. Our study is the validity of the method to replace LD with linkage in association study, for which we investigate the statistical power of different heritability and the number of QTL using simulation data. We found that it is entirely feasible way to exploit the multiple regression method for Genome-wide association study using only linkage information. Similar to the typical genome wide association tests using LD information, our new approach performed validly when the multiple regression based on linkage was employed. However, the performance improved slightly when the linkage is used alone, which was much closer to the traditional GWAS model using single marker regression. Meanwhile, the results show that the statistical power of the new method decreases with the increase of the number of QTL, and its power is sensitive to heritability. To sum up, these results suggest that it can identify QTL, although the power is relatively weak. It remains to be discovered what really causes this phenomenon.

100

Identification of genomic islands from sequence data in Atlantic cod (Gadus morhua L.)

Dr. Silvia Teresa Rodriguez Ramilo[1], M. Baranski[2], H. Moghadam[2], H. Grove[3], T. Graceline Kirubakaran[3], S. Lien[3], T. H. E. Meuwissen[3], A. K. Sonesson[2]
[1]INRA, [2]Nofima AS, [3]Department of Animal and Aquacultural Sciences, University of Life Sciences

Low levels of coancestry and weak population genetic differentiation are expected in Atlantic cod. The reason is that they exploit a wide geographical distribution with different ranges of salinity and temperature. However, two ecologically distinct ecotypes have been described: Northeast Arctic cod (NEAC) and coastal cod (CC). The objective of the present study was to identify genomic regions showing a differential behavior between NEAC and CC individuals. Genomic information from 93 animals (9 CC individuals, 50 NEAC animals and 34 CC × NEAC crossed individuals) and 3,123,434 autosomal SNPs were available for the study. Genome-based coancestry, genetic differentiation coefficient, private alleles and principal component analysis tests were performed. Genome-based coancestry estimates and the unique allele variants content were calculated following a sliding windows approach of 200 SNPs. Divergent genomic regions were identified on chromosomes 1, 2, 7 and 12, extending approximately 18, 7, 11 and 13 Mb, respectively. These results indicate that both ecotypes are diverging in four chromosomes, while the majority of the genome remains relatively homogenized.

101

Identification of the causal mutation for transverse hemimelia in River Buffalo using whole genome sequence data

Lynsey Whitacre[1], Jesse Hoff[1], Robert Schnabel[1], Sara Albarella[2], Francesco Strozzi[3], Chiara Ferrandi[3], John Williams[4], Jeremy Taylor[1], Jared Decker[1]
[1]University of Missouri, [2]University of Federico II – Naples, [3]Parco Tecnologico Padano, [4]University of Adelaide

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 river buffalo populations and causes significant production loss in affected animals. Further, carriers of the disease 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 11 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 approximately 19 million single nucleotide polymorphisms (SNPs). Based on case versus control SNP variant analysis, genome-wide association studies (GWAS), and homozygous mapping via whole genome de novo assembly, several candidate variants have been identified. However, candidate regions or genes have not yet been identified because the buffalo reference sequence has not been place on chromosomes and the relative locations of variants associated with the disease are unknown. Ongoing research aims to define candidate regions by aligning buffalo scaffolds with significant association to the cow reference genome, which is far more complete in its structure and annotation.

102

Candidate variants confirmation by GWAS applied to large cow data sets

Mr. Andrew Marete[1,2,3], Dr Didier Boichard[1], Prof Mogens S. Lund[2]
[1]INRA, [2]Center for Quantitative Genetics and Genomics, Aarhus University, [3]AgroParisTech

This research aims to validate putative candidate mutations in three French dairy breeds. The data used included 79,532 genotyped cows with full pedigrees and 42 phenotypes including various trait classes based on production, morphology, fertility and clinical mastitis.
The animals were genotyped either with the 50K chip or with Eurogenomics custom chip (Euro10kG), which contains 17,232 SNPs. Of these, 7,232 SNPs are custom selected and contain SNPs believed to be highly informative to genomic selection whereas the remaining are generic markers from the Illumina BovineSNP50® beadchip. Imputation to 50K was then done using the FImpute software.
The validation was twofold and done on the imputed genotypes. First association studies were carried out to analyze all SNPs separately. A Genome Wide Association Study (GWAS) model was fitted per breed, per chromosome and per trait and analysis done using GCTA software. Secondly, using GS3 software, all markers were analyzed simultaneously using a Bayesian approach (BaysCπ) with a chosen prior of 0.001 and including genotypes, phenotypes and pedigree information.
Results of the fitted GWAS model showed that 172 SNPs from the custom chip were significantly better in fifteen traits compared to SNPs from the 50K ($P<0.001$). Comparison of SNPs within same traits and between breeds showed that a total of 6,386 SNPs were highly associated ($P<0.001$). The stringent prior of the Bayesian model was aimed at reducing noise in the model and allowing only the best SNPs to be selected on each iteration. The results were consistent with GWAS model such that 138 SNPs from the custom chip had higher probabilities of inclusion compared to 50K markers (Bayes Factor>130).Meta-analysis using METAL software was done to consolidate the two results and a validated SNP list generated.

103

Inclusion of non-additive effects on G-BLUP model considering different kinship matrices

Mr. Indalécio Vieira[1], Mr. Jhonathan Pedroso dos Santos[2], Mr. Luiz Paulo Miranda Pires[1], Mrs. Flávia Avelar[1], Mr. Márcio Balestre[1]
[1]UFLA, [2]ESALQ

The comprehension of non-additive genetics effects in the expression of quantitative characters is very important during the genotype selection phase, specially for species in which the commercial products are clone or hybrids. The use of molecular markers has enabled the investigation of these effects at the genomic level and a better comprehension of its importance for quantitative traits. Nowadays additive genetic effects are usually predicted using G-BLUP (Genomic Best Linear Unbiased Predictor) model, however, inclusion of non-additive effects are exception. Therefore, the aim of this work was to evaluate the behavior of the G-BLUP model incorporating different genetic effects and kinship matrices. For this purpose, data of the circumference at breast height (CBH) in Eucalyptus spp was analyzed using three kinship matrices proposed by the authors: VanRaden (V), Meuwissen (M) and Yang (Y) combined with three different models: reduced, considering only additive effects; additive-dominant which consider additive and dominance effects; and the full model, incorporating additive, dominance and epistatic effects. It was estimated variance components and their standard errors and the heritability considering each model and kinship matrix to verify the parameter estimate efficiency. There was prevalence of additive genetic effects for all models and kinship matrices. The additive genetic effects were estimated with good precision in all cases, on the other hand non-additive effects had a poor estimate accuracy. The matrices M and Y had similar behavior, that is, they basically recovered the same information. It is possible to conclude that models which handle better with lack of orthogonality of non-additive effects are needed, and there isn't expected significant differences on results choosing matrix M or Y. In addition, it can be inferred that the trait CBH is mainly controlled by additive genetic effects.

104

Marker-based estimates reveal significant non-additive effects in clonally propagated cassava (Manihot esculenta): implications for the prediction of total genetic value and the selection of varieties

Dr. Marnin Wolfe[1], Dr. Peter Kulakow[2], Dr. Ismail Y. Rabbi[2], Dr. Jean-Luc Jannink[1,3]
[1]*Cornell University,* [2]*International Institute of Tropical Agriculture,* [3]*USDA-ARS*

In clonally propagated crops, non-additive genetic effects can be effectively exploited by the identification of superior genetic individuals as varieties. Cassava (Manihot esculenta Crantz) is a clonally propagated staple food crop that feeds hundreds of millions. We quantified the amount and nature of non-additive genetic variation for three key traits in a breeding population of cassava from sub-Saharan Africa using additive and non-additive genome-wide marker-based relationship matrices. We then assessed the accuracy of genomic prediction of additive compared to total (additive plus non-additive) genetic value. We confirmed previous findings based on diallel populations, that non-additive genetic variation is significant, especially for root yield. Further, we show that total genetic value predicts observed phenotypes more accurately than additive value for root yield but not for dry matter content, which is mostly additive or for CMD resistance, which has high narrow-sense heritability. We address the implication of these results for cassava breeding and put our work in the context of previous results in cassava, and other plant and animal species.

105

Leveraging the use of genomic resources to guide genomic selection

Dr. Dunia Pino Del Carpio[1], Dr Marnin Wolfe[1], Dr Jean-Luc Jannink[1]
[1]*Cornell University*

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.
In the present study we used genome-wide association results of traits with different genetic architecture as weights when constructing the genomic matrix (GRM) for genomic prediction. Weighting factors of the SNPs in the genomic relationship matrix were the squared value of the SNP effect, and the corresponding P-value [−log10(P)]. Additionally, statistical models can be improved if structural genome annotation is included partitioning SNPs based on their annotation. In studies from human genetics classification of SNPs according to their genomic location in genes (coding, promoter, UTR, intron), or intergenic locations has shown differentially enrichment of variance explained.
In the present study we followed a GBLUP multi-kernel approach partitioning SNPs based on their genome functional annotation. We first estimated the heritability value of each functional kernel using an LD correction. This heritability value was used to calculate an enrichment term that was used together with GWAS P-value [−log10(P)] to estimate GRM weights for genomic prediction. Finally, we evaluated the use of biological data with SNPs annotation to pathways; prediction was carried out using Random Forest and Mixed Random Forest.
Results and implications of the study in terms of the genomic prediction accuracies are discussed.

106

Pigmentation and pupation height phenotypic plasticity catalyzed by desiccation in worldwide populations of Drosophila melanogaster

Ms. Audrey Simard[1], Dr. Heloise Bastide[1], Dr. John Pool[1]
[1]*University of Wisconsin-Madison*

Drosophila melanogaster has colonized a range of varying climates in Sub-Saharan Africa and beyond. The resulting local adaptation provides an ideal model system for studying the genetic architecture of adaptive trait evolution. Here, we quantified desiccation tolerance and pupation height diversity among populations of Drosophila melanogaster. We also investigated the effects of desiccation exposure during early development stages to adulthood by observing the phenotypic plasticity of pigmentation and pupation height under humid versus dry conditions. Desiccation tolerance varied substantially among populations, reaching its maximum in an Ethiopian population adapted to survive in high altitudes. While known to vary among our populations, abdominal pigmentation showed no clear plastic response to desiccation. Pupation height was also found to vary among populations of Sub Saharan African, and cosmopolitan populations such as Peru and France. This trait showed population-dependent phenotypic plasticity, in populations vulnerable to desiccation stress showed much lower pupation heights in drier conditions, while the Ethiopian population with the highest desiccation tolerance had the least plasticity in pupation height. These findings clarify relationships among important phenotypes in D. melanogaster, and identify desiccation tolerance variability as a potential target for genetic analysis.

107

Rare variants explain a substantial fraction of the genetic variance for female fertility in dairy cattle

Dr. Goutam Sahana[1], Ms. Qianqian Zhang[1], Dr. Bernt Guldbrandtsen[1], Prof. Mogens S Lund[1]
[1]*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University*

Quantifying the relative contributions of rare and common variants to trait variation is important to disentangle the genetic architecture of quantitative traits and complex diseases. Theoretical and empirical studies suggest that rare variants (defined as those have minor allele frequency (MAF) < 0.01), could play a significant role in quantitative trait variation. We studied the contribution of rare genetic variants to the total genetic variance for fertility, a low heritability trait in dairy cattle. The response variable used was de-regressed breeding values (DRPs) for fertility index. About 5,000 Holstein bulls genotyped with 50k SNP array were imputed to full genome sequence variants using a two-step approach. 50k genotypes were first imputed to a high-density SNP array (777k) using a multi-breed reference of 3,383 animals followed by imputation to whole genome sequence variants using 1,228 whole genome sequences as reference. The number of genetic variants in the imputed whole genome sequence after quality control was ~15 million. We used a MAF-stratified GREML approach (GREML-MS) as implemented in GCTA software. The markers were grouped into three classes based on MAF (>0.1, 0.1-0.3 and >0.3). The total variances of DRPs for fertility index explained by all the sequence variants were 60.9%. The SNPs with MAF < 0.01 accounted for 7.6%; in comparison, common SNP variants (MAF>0.01) explained 53.2% of the total variance for DPRs for fertility index. These results suggest that rare variants explain a substantial percentage of the total genetic variance for female fertility in cattle and therefore can be exploited to increase accuracy in genome-wide prediction.

108

Recombination and genetic variance among maize doubled haploids induced from F1 and F2 plants

Joshua Sleper[1], Rex Bernardo[2]
[1]*Syngenta Seeds,* [2]*University of Minnesota - Twin Cities*

Maize (Zea mays L.) breeders rely on doubled haploid (DH) technology for fast and efficient production of inbred lines. Breeders can induce DH lines most quickly from F1 plants (DHF1), or induce DH lines from F2 plants (DHF2) to allow selection prior to DH induction and have more recombinations. Our objective was to determine if the additional recombinations in maize DHF2 lines leads to a larger genetic variance and a superior mean of the best lines. A total of 311 DHF1 and 241 DHF2 lines, derived from the same biparental cross, were crossed to two testers and evaluated in multilocation trials in European and the U.S. The mean number of recombinations per genome was 14.48 among the DHF1 lines and 21.38 among the DHF1 lines. The means of the DHF1 and DHF2 lines did not differ for yield, moisture, and plant height. The genetic variance was higher among DHF2 lines than among DHF1 lines for moisture, but not for yield and plant height. However, the higher genetic variance for moisture among DHF2 lines did not lead to a lower moisture of the best 10% of the lines. Our results indicated that the decision of whether to induce DH lines from F1 or F2 plants needs to be made from considerations other than the performance of the resulting DHF1 or DHF2 lines.

109

Genomic prediction based on interaction networks of markers: How to incorporate prior experimental information

Dr. Johannes WR Martini[1], Dr. Valentin Wimmer[2], Prof. Dr. Henner Simianer[1]
[1]*Georg-August University,* [2]*KWS SAAT SE*

Different approaches have been developed in quantitative genetics to model non-additive relationships of individuals. Among them, the epistasis marker effect model which assumes that the genetic value is not only the sum of additive effects, but additionally of pairwise interactions of genetic markers, has been discussed in the last years. Dealing with the full epistasis model with all pairwise interactions, two characteristics attract attention. One obvious fact is that the number of variables in the model increases drastically from the number of markers p in the purely additive model to approximately p^2/2 in the full epistasis model. Since the model also incorporates a large number of potentially unimportant variables, the question arises whether predictive ability can benefit from variable selection. Another important point is the coding-dependent performance of the epistasis model in genomic prediction, raising the question of how to code the markers optimally for a given data set. We show that both characteristics offer possibilities for incorporating knowledge from previous experiments into the prediction method: Instead of modeling all pairwise interactions, prior knowledge from previous experiments can be used to infer interaction subnetworks by restricting the model to biologically relevant interactions. These subnetworks which may very roughly reflect genetic architecture of the trait can afterwards be used for other data sets as prior information. Moreover, the coding-dependent performance of the epistasis model allows for adapting the coding to previous results which can afterwards be used as prior information in new experiments. The presented results illustrate that properties of the epistasis model which may at first seem unattractive can also provide the flexibility to reflect genetic architecture in at least two ways.

110

Whole genome hitchhiking on an organelle mutation

Dr. Padraic Flood[1,2,3], Dr. Joost van Heerwaarden[2], Dr. Jeremy Harbinson[2], Dr. Mark G.M. Aarts[2], Mr. Frank Becker[2], Mr. C. Bastiaan de Snoo[3]
[1]Max Planck Inst for Plant Breeding Research, [2]Wageningen University, [3]Rijk Zwaan R&D

Strong selection on a beneficial mutation can cause a selective sweep, which fixes the mutation in the population, and reduces the genetic variation in the region flanking the mutation. These flanking regions have increased in frequency due to their physical association with the selected loci, a phenomenon which is called 'genetic hitchhiking'. Theoretically, selection could extend the hitchhiking to unlinked parts of the genome, to the point that selection on organelles affects nuclear genome diversity. Such indirect selective sweeps have never been observed in nature. Here we show that strong selection on a chloroplast gene in the wild plant species Arabidopsis thaliana, has caused widespread and lasting hitchhiking of the whole nuclear genome. The selected allele spread more than 400 km along the British railway network, reshaping the genetic composition of local populations. This demonstrates that selection on organelle genomes can significantly reduce nuclear genetic diversity in natural populations. We expect that organelle mediated genetic draft is a more common occurrence than previously realised and needs to be considered when studying genome evolution.

112

Epigenome-wide association study and prediction of triglyceride postprandial responses to high-fat dietary challenge

Dr. Chao-Qiang Lai[1], Dr. Mary K. Wojczynski[2], Dr. Bertha Hidalgo[3], Dr. Laurence D. Parnell[1], Dr. Marguerite Ryan Irvin[3], Dr. Stella Aslibekyan[3], Prof. Michael A. Province[2], Dr. Devin Absher[4], Prof. Donna K. Arnett[3], Prof. Jose M. Ordovas[1]

[1]USDA ARS HNRCA At Tufts University, [2]Department of Genetics, Washington University School of Medicine, [3]Department of Epidemiology, School of Public Health, University of Alabama , [4]Hudson Alpha Institute for Biotechnology

The magnitude of postprandial lipemia (PPL), the increased plasma triglyceride (TG) concentration resulting after consumption of a dietary fat meal, appears to be an independent risk factor of cardiovascular disease (CVD). Therefore, its prediction and management could lead to effective strategies for prevention of CVD. Individual's responses to a fat meal vary greatly, depending on genetic and lifestyle factors. However, only a few loci have been associated with TG PPL response and most of the genetic variation remains unexplained. Heritable epigenomic changes (i.e., DNA methylation) may be significant contributors to the unexplained inter-individual PPL variability in response. We conducted an epigenome-wide association study (EWAS) with 979 subjects who were challenged with a high-fat meal (83% calories from fat, 14% from carbohydrate, and 3% from protein) in the Genetics of Lipid Lowering Drugs and Diet Network study. DNA methylation patterns were assessed using Illumina Human 450K methylation array. EWAS was conducted using a linear mixed model adjusting for sex, age, study site, and the first four principal components for T-cell impurity as fixed effects, and kinship based on family relationship as a random effect. Eight methylation sites, encompassing five genes: LPP, CPT1A, APOA5, SREBF1, ABCG1 were significantly associated with PPL response at epigenome-wide level (P<10-7). Higher methylation status at LPP, APOA5, SREBF1, ABCG1 was correlated with increased TG PPL response, whereas lower CPT1A methylation level was linked to increased TG PPL response. These methylation sites account for a substantially greater amount of phenotypic variance (22.7%) of PPL when compared to genetic contribution of loci identified by our previous genome-wide association study (5%). In summary, the epigenome has a large contribution to the variation in PPL and this may be used in the future to modulate PPL to reduce CVD.

113

Bayesian inference of the allelic series at quantitative trait loci in multiparent populations

Wesley Crouse[1], William Valdar[1], Samir Kelada[1]
[1]*University of North Carolina-Chapel Hill*

Multiparent populations (MPPs) such as the Collaborative Cross (CC, Mus musculus), the DSPR (Drosophila melanogaster), and the MAGIC lines (Arabidopsis thaliana) have become increasingly popular resources in quantitative genetics. MPPs are generated by outcrossing diverse inbred founder lines, yielding a population in which every individual genome is mosaic of the original founder haplotypes. Such populations provide distinct advantages for detecting quantitative trait loci (QTL) because tests of association between phenotypes and alleles can leverage inferred haplotype descent. Once a QTL is detected, however, further analysis is required to determine how inferred haplotypes group into distinct alleles, i.e., the allelic series or “strain distribution pattern” (SDP). We propose a Bayesian framework for inferring the SDP that takes into account sources of uncertainty found in typical MPPs, including individuals’ haplotype state at the QTL, the number of functional alleles, and the magnitude of their effects.

In applying our method to simulated and real data from incipient CC lines, we consider two main SDP priors: a uniform prior on the space of alternative groupings and an informative prior based on mutation rates and local phylogeny. The uniform prior results in low posterior certainty on any one SDP but permits high-confidence allelic contrasts and discriminatory scoring of candidate causal variants. The informative prior is powerful in identifying SDP that are consistent with local phylogeny but is sensitive to misspecification of the causal region. We discuss how uncertainty about both location and phylogeny can be modeled through our framework, and also the utility of Dirichlet and similar process priors for this kind of inference. Our method can be generalized to multiple phenotypes — we show how joint estimation of the SDP shared across phenotypes increases posterior precision — and is applicable to other MPPs, including those with a large number of founder lines.

114

Genomic data integration for GWAS and eQTL analysis

Constanza Rojo[1], Qi Zhang[2], Sunduz Keles[1]
[1]*University of Wisconsin-Madison,* [2]*University of Nebraska Lincoln*

Genome wide association studies (GWAS) are widely used to elucidate genetic variation associated with traits. When combined with genome-wide expression analysis (eQTL), these studies have the potential to identify genetic variants that modulate disease or trait phenotypes through their impact on expression of genes. Challenging aspects of these studies are: high dimensionality of the variant space and that most candidate genetic variants are located either in intronic or intergenic regions making their interpretation challenging. We develop a novel mediation analysis framework to integrate eQTL and GWAS studies with the effective utilization of publicly available annotation data. This approach extends the scope of standard mediation analysis framework adapted by statistical genetics from accommodating 10s of genetic variants to 100s to 1000s of variants. The integrative framework incorporates the regulatory information encoded in annotation data to elucidate how top-ranking association SNPs might be modulating gene expression. We apply our method to Framingham Heart Study data of type 2 diabetes and identify expression modulating non-genic genetic variants with significant impact on diabetes.

115

Multivariate genome-wide association study (GWAS) for detailed milk protein composition

Mr. Grum Gebreyesus[1,2], Dr M.S. Lund[1], Dr L Janss[1], Dr A.J Buitenhuis[1]
[1]*Aarhus University,* [2]*Wageningen University*

Studies based on simulation have shown that multivariate Quantitative trait loci (QTL) mapping models have an increased power to map QTL compared to univariate GWAS. This might be of special interest in the case of scarcely recorded traits, such as milk protein compositions, that are only measured on experimental animals due to requirement of costly equipment and time taking measurements. This study aimed at assessing, based on real data, the advantages of multivariate GWAS for scarcely recorded protein composition making use of additional information from routinely recorded traits. Studied traits included protein percentage, αS1-CN, αS2-CN, β-CN, κ-CN, α-LA, β-LG, glycosylated-κ-CN and αS1-CN-8P measured on 650 Holstein cows. The multivariate GWAS models were run with each detailed protein fraction analyzed in combination with de-regressed proofs (DRPs) in milk protein yield obtained for 5200 bulls. Non-existing residual covariance were set as each pair of traits in the multi-trait analysis were measured on different individuals. The polygenic random effect was added to the model using the additive pedigree relationship matrix to correct for possible family structure effects and fitting a SNP at a time independently. Alternatively, univariate GWAS was run for each trait to compare the performance of the multi-trait models. Significance thresholds were determined using a false discovery rate (FDR) correction. Model fit was assessed with AIC values. No significant SNP markers were detected for αS1-CN, α-LA and αS1-CN-8P using the standard single-trait GWAS. In comparison, 3 markers were found significantly associated for α-LA using multivariate GWAS. More markers were found to have significant association for glycosylated-K-CN (+26) and K-CN (+50) using multivariate GWAS than univariate analysis. The multivariate models also had the lowest average AIC value. These results could be confirmation, based on real data, of the advantages of multivariate GWAS models making use of pseudo-observations for scarcely recorded traits.

116

VALIDATE on the CyVerse cyberinfrastructure: scalable testing of the accuracy and precision of association and prediction software

Ann Stapleton[1], Danielle Chaung[1], K. Sierra Cockerill[1], Stephen P. Talley[1], Samuel W. Buck[1], Mitch Powell[1], James S. Gray[1], Geoffrey Carpenter[1], Samuel Capeloto[1], Dustin Landers[1], CyVerse[2]
[1]UNCW, [2]CyVerse.org

Genotype and phenotype datasets (SNP, imaging, geospatial, etc.) often exceed the abilities of researchers to share and analyze. Constantly changing software and technologies compound challenges that concern scientists and core facilities managers as they scale to meet these needs, including the rise of interdisciplinary and inter-institutional collaborations.
CyVerse is a US National Science Foundation (NSF) funded cyberinfrastructure project to address these challenges. Web-accessible tools and well described application interfaces for data analysis and management of data-driven collaborations leverage federated data and consumption of resources from multiple providers such as NSF funded XSEDE (eXtreme Science and Engineering Discovery Environment), campus clusters, and commercial clouds. Researchers can securely manage and share their data, software tools, and analysis pipelines with their collaborators and/or a large community of users without provisioning their own underlying computational infrastructure.
We have constructed a publically accessible CyVerse data analysis workflow, VALIDATE, that makes it easier to test the accuracy and precision of genotype-to-phenotype association and prediction applications. This is done using existing or your own custom known-truth simulations -- which have large numbers of files or large numbers of iterations and thus substantial computational demand.
Our Winnow app is used to evaluate the other tools in your pipeline. Given the "known truth" of a simulation data set, we output a series of fit statistics to determine how a method was useful in analyzing specific datasets. We support a range of simulation software and stored files, with several pre-installed association and prediction applications. Since the analysis are run on the XSEDE computer systems, the computational demand for your simulations will scale up as your needs increase. We especially encourage you to make your new apps public and share them through CyVerse, so that people who adopt your method can run it quickly and easily.

117

Contemporary group estimates with the inclusion of minimum monthly temperature in the model increases the ability to detect genotype by environment interactions in pigs

Mrs. Sarita Z.Y. Guy[1], Associate Professor Susanne Hermesch, Associate Professor Peter C. Thomson
[1]*Faculty of Veterinary Science, The University of Sydney,* [2]*Animal Genetics and Breeding Unit, a joint venture of NSW Department of Primary Industries and University of New England, The University of New England*

Estimates of contemporary group (CG) effects are commonly used to quantify the known and unknown influences of the environment that pigs experience. This study investigates the inclusion of minimum air temperature in the models to estimate CG effects, which were then used to analyse genotype by environment interaction (G×E) for growth.
The data were recorded from 1999 to 2013 at a piggery in south-east Queensland, Australia (n=34,116). The Australian Bureau of Meteorology website provided monthly averages of daily minimum air temperatures (MinTemp).
For the first stage of analysis, CG effects were estimated for average daily gain, recorded at an average body weight of 91 kg. The base model contained the fixed effect of sex, and random effects of polygenic animal effect, litter effect and birth-month-year CG. The second model included the effects of the base model and MinTemp of the test month as an additional fixed effect. Estimates of CG from each model were portioned into four groups, which were then used as the environmental variables in the second stage of analysis. Two sire × environment interaction models were fitted containing the same fixed effects as the models in stage one, but with the random effects of sire, litter effect, CG, and sire × environment.
As expected, the estimated variance of CG effects declined from 543.6±62.9 to 282.6±36.3 with the inclusion of MinTemp in the first stage of analysis. Other variance components did not change appreciably between the models.
The inclusion of MinTemp in the models of both steps increased the sire × environment variance, from 22.3±13.0 to 79.4±19.9. Although explaining only 1.5% of the phenotypic variation, the ability to detect G×E in a high health herd increased. Therefore, the inclusion of temperature in the model provides a finer definition of hard-to-measure or unknown environmental aspects, including possible health challenges.

118

Identifying recombination events and haplotypes in beef cattle from half-sib families

Mr Mohammad Hossein Ferdosi[1], Dr Vinzent Börner[1], Prof Bruce Tier[1]
[1]*Animal Genetics and Breeding Unit, University of New England*

Identifying the recombination events and haplotype segments that have been passed to offspring is useful for several applications in genomic analysis. In livestock populations half-sib families are common, therefore development of methods for identifying recombination events and phasing in half-sib families will allow us to phase most individuals in the population. The generated haplotypes make the phasing of the remained individuals straight forward. Subsequently haplotypes can be used to build the genomic relationship matrix. In this study we propose a likelihood based method that is more flexible than the rule based methods for development and to handle incorrect genotypes. This algorithm first imputes sire haplotypes from offspring genotypes and then creates a descent graph from parent to offspring that shows the recombination events in the sire and origin of the paternal haplotype in offspring. With knowledge of recombination events and haplotype segments passed to the offering, phasing the haplotypes of offspring become feasible and simple. In addition to the phasing and imputation, this method had several benefits such as identifying pedigree errors and mapping errors. The method was tested on real genotypes of 14,600 beef cattle. The accuracy of sire imputation was ≥ 95% with the average of 98% for the half sib family size of 7 to 79. The average number of recombination events in paternal haplotype strands transferred to the offspring was 26.5 recombinations with a standard deviation of 6.3. In conclusion, the high sire imputation accuracy indicated a high accuracy in haplotype inference. This method can remove the need for sire genotyping and allow us to identify unusual recombination events that may indicate mapping or pedigree errors.

119

Linkage drag connects spike and root development in wheat

Mr. Kai Voss-Fels[1], Mr. Lunwen Qian[1], Mr. Sebastian Parra-Londono[2], Mr. Ralf Uptmoor[2], Mr. Matthias Frisch[3], Mr. Gabriel Keeble-Gagnère[4], Mr. Rudi Appels[4], Mr. Rod Snowdon[1]
[1]Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Germany, [2]Department of Agronomy, University of Rostock, Germany, [3]Institute for Agronomy and Plant Breeding II, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University Giessen, Germany, [4]State Agriculture Biotechnology Centre, School of Veterinary & Life Sciences, Murdoch University, Western Australia

Improving the roots of modern bread wheat varieties is considered to be a key factor in breaking yield barriers and improving adaptation to abiotic stress in the face of climate change. However, the plants' "hidden half" has been largely ignored to date. Effective genome-based selection for improved root systems would add an important tool to the breeders' toolbox.

We performed comprehensive greenhouse phenotyping for basic above- and below-ground growth traits in an extremely diverse collection of 215 hexaploid wheat accessions from Europe, China, Australia and North America. Genome-wide association mapping and haplotype network analysis using high-density single-nucleotide-polymorphism (SNP) markers identified rare genetic variants that are significantly associated with root dry mass at seedling and adult plant stage. These loci lie within a highly conserved chromosomal block with exceptionally high linkage disequilibrium, suggesting that this region has been under strong directional selection. Subsequent bioinformatics analysis and sequential alignment identified adjacent candidate genes for both spike and root development, indicating inadvertent co-selection of specific root variants during breeding for heading date. Modelling of epistasis within this chromosomal block further revealed extremely strong interactions between and among loci associated with root and spike development.

In summary, our results provide breeders the possibility for a genomics-based exchange of co-localized haplotypes to maximize root variation, while maintaining specific heading characters. They furthermore allow insights into non-pleiotropic interactions of loci associated with root and spike architecture, and enable the reversal of unintended effects on genetic variability due to linkage drag. Ongoing work based on our findings will help to functionally validate these novel uncharacterized candidate genes with a major effect on underground plant development.

120

Using metafounders reduces bias in single step GBLUP genomic evaluations

Carolina Garcia-Baccino[1], Zulma G. Vitezica[2], Andres Legarra[2], Rodolfo J.C. Cantet[1]
[1]*Universidad de Buenos Aires, CONICET,* [2]*INRA*

Combining genomic and pedigree relationships requires the same base populations, that is, same average breeding values and genetic variance at the base. This is difficult to do when genotyped animals are younger or selected. To make base populations identical we consider each ancestral population as a finite-sized pool of gametes called metafounder. A metafounder is a pseudo-individual included as founder of the pedigree and similar to an "unknown parent group".
The use of metafounders enables the modification of pedigree relationships to match genomic relationships. The latter are computed using the -101 coding for genotypes, then the relationship of the metafounder (gamma) is estimated using pedigree and markers. Gamma is proportional to the (co)variance of allelic frequencies in the pedigree founders. Allele frequencies were estimated using least squares (LS), generalized least squares (GLS) or maximum likelihood (ML).
We carried out a simulation study (20 replicates) to explore the possibility of including metafounders in Single Step GBLUP (SSGBLUP). Data was simulated with QMSim mimicking a dairy cattle selection scheme. Ten recent generations were simulated coming from 100 historical generations including 40,000 segregating markers and 1,500 segregating QTLs. There were 28,800 individuals in the pedigree, 6,038 individuals with genotypes and only the females had phenotypic records. Generation 10 had no phenotype and was used for checking accuracy and bias.
Estimates of gamma were accurate and unbiased if GLS or ML were used (0.40 and 0.39, respectively; 0.40 was the true value) but overestimated with LS (0.43). There was an important gain in terms of bias when including a metafounder in SSGBLUP (the regression coefficient of true on estimated breeding values was 0.82 for regular SSGBLUP and 0.94 when including the metafounder). There was also a slight gain in accuracy (0.72 vs 0.74). Using metafounders also produced more accurate variance components estimations.

121

Genetic analysis of productive life in Chinese Holstein population

Xinyi Yan[1], Aoxing Liu[1,2], Ganghui Dong[3], Gang Guo[3], Xizhi Li[3], Guosheng Su[2], Yachun Wang[1]
[1]College of Animal Science and Technology, China Agricultural University, [2]Center for Quantitative Genetics and Genomics, Aarhus University, [3]Beijing Sunlon Livestock Development Co., Ltd.

(Objective) Productive life (PL) or longevity is an economically important trait in dairy cattle. However, genetic evaluation of longevity in China has not yet been performed. The objectives of this study were to establish the model and estimate the genetic parameter for PL, which will facilitate the selection program related to longevity in China in the future.
(Method) The data set consisted of culling records of 47, 878 Chinese Holstein cows with calving years during May 1996 to January 2012 from 32 herds in Beijing, China. These cows had complete records with regard to date of birth, culling, and first calving. PL was defined as the total days from first calving to culling. Pedigree was traced back to minimum 3 generations and the unknown parents were assigned into 16 phantom groups based on sex and birth year. A linear mixed model, including fixed effects of age at first calving, herd-birth year, birth year-month and random individual additive genetic effect was used to estimate variance components and breeding values using the DMU package.
(Result) The results showed that the average PL in Chinese Holstein cows was about 2.7 years, the range of PL from different birth year was 786~1051 days. The regression coefficient of age at first calving (in month) was -25.1±0.778. There was no clear phenotypic trend of PL from this data set when birth year or culling year was considered. The estimated heritability of PL was 0.083±0.010. Average accuracy of estimated breeding values for sires with minimum 20 daughters was 0.768.
(Conclusion) Longevity is a low heritability trait, greatly influenced by environmental factors. However, with progeny test or with genomic prediction, the genetic improvement of PL through selection is feasible, and balanced breeding strategies can be adopted.
Key words: Chinese Holsteins; productive life; heritability; estimated breeding value

122

Genetic parameter estimation in purebred and crossbred pigs including dominance with kernel and two genomic models

Dr. Llibertat Tusell Palomero[1], Dr. Andrés Legarra[1], Dr. William Herring[2], Dr. Zulma Vitezica[1]
[1]INRA, INPT, UMR GenPhySE, F31326 Castanet-Tolosan, France, [2]PIC, Hendersonville, USA

Recently, we proposed an approach to estimate genetic variance components in an F1 population under a genomic model with additive and dominant gene action (GBLUP-F1). Using GBLUP-F1, genetic variance is partitioned into additive variance (due to substitution effects of the parental gametes) and dominance deviations. Semi-parametric models using kernels (RKHS) to account for dominance may be more flexible and powerful. This work compares variance components estimated with kernel matrices, with "raw" genomic matrices and with the GBLUP-F1.

The genomic matrices involved biological additive (-1/0/1 for the AA/Aa/aa) and dominant (0/1/0) effects of genes. The RKHS model used two Gaussian kernel matrices based on Euclidean distances involving either gene content (additive kernel) or genotypes (dominant kernel). The different variance components were estimated by REML under each model and then referred to the same base population multiplying them by Dk=avg(K)-avg(diag(K)), being K the kernel or the "raw" genomic matrices used. We analyzed litter size in pig data provided by Genus plc (Hendersonville, TN, USA) including purebred populations and their crossbreds.

After the correction with Dk, the total genetic variances estimated with kernel and genomic matrices were compared with those obtained with GBLUP-F1; they were identical in all models. Partition of the variance in dominance and additive components was, however, different between the kernel/genomic models and the GBLUP-F1. The partitions using kernel or "raw" genomic matrices are not orthogonal. Semi-parametric models using kernels (RKHS) may be more flexible and powerful, but the results that they produce are much more difficult to interpret. GBLUP-F1 models perform more satisfactorily to explain the partition of the total genetic variance.

123

Distinct genetic architectures for maize phenotypes and phenotypic plasticity

Aaron Kusmec[1], Dan Nettleton[2], Patrick S. Schnable[1]
[1]*Department of Agronomy, Iowa State University,* [2]*Department of Statistics, Iowa State University*

Phenotypic plasticity describes the phenotypic variation of a trait when a genotype is exposed to multiple environments. As world population growth and climate change stress food production systems, understanding the genetic control of phenotypic plasticity of crops such as maize becomes of paramount importance. Here, we report the results of genome-wide association analyses of multiple phenotypes and two measures of phenotypic plasticity in the maize nested association mapping (NAM) populations grown in multiple environments and genotyped with ~2.5 million single nucleotide polymorphisms. We show across all measured traits that main effect and plasticity candidate genes form distinct groups. Within each phenotype, candidate genes for main effects and plasticity are enriched for different but complementary functions. Such independent genetic control of main effects and plasticity suggests that breeders could independently select for optimum degrees of performance and plasticity, thereby exploiting plasticity according to the demands of the target performance environment.

125

Recombination patterns among spring barley populations

Dr. Ana M. Poets[1], Alexandrea Ollhoff[1], Kevin P. Smith[1]
[1]*University of Minnesota*

The number and location of recombination events vary among individuals. The most commonly used methods to identify the genes underlying traits assume that recombination frequencies are consistent across families. Additionally, these methods often suffer from the presence of too little (e.g. in QTL mapping) or too much recombination (e.g. in association mapping). Nested association mapping (NAM) populations help resolve the issue by combining the power of detection of a QTL mapping population with the resolution gain from more recombination in an association mapping population. We used a six-row spring barley NAM population of 7,700 RILs representing 91 families to characterize recombination in spring barley. A combination of exome capture and genotyping-by-sequencing technology were used to collect dense genome-wide single-nucleotide polymorphism data. These data were used to develop a dense consensus genetic map and examine variation in fine scale recombination patterns among barley families. This high-resolution genetic map will be used to relate recombination to physical distance using the most recent barley reference genome.

126

Polymorphisms in PDK4 and DMRT3 genes in Arabian and Quarter Horses

Inaê C Regatieri[1], Guilherme L Pereira[2], Rogério A Curi[2], Guilherme C Ferraz[1], Antonio Queiroz-Neto[1]
[1]Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, [2]Faculdade de Medicina Veterinária e Zootecnia, UNESP – Univ Estadual Paulista

The PDK4 gene is directly involved in the process of ATP production, and the G allele of the PDK4 SNP (single nucleotide polymorphism) stimulates the beta-oxidation of fatty acids, an important pathway for energy production in equines that perform long and submaximal intensity exercises, like Arabian horses. The DMRT3 gene is involved in the coordination of the locomotor system that controls movement and is associated with gait in some breeds of horse. This study investigated polymorphisms of PDK4, and DMRT3 genes in Arabian and Quarter Horses in order to associate polymorphisms with performance. Arabian horses were divided into high and low performance groups based on success in endurance competition, whereas racing Quarter Horses were separated by a speed index. Polymorphisms of the genes of interest were analysed by ARMS-PCR (PDK4) and PCR - RFLP (DMRT3) techniques. To compare the frequencies of polymorphic SNPs between the groups, the Fisher's exact test was performed in the software R with significance level of 5%. For the DMRT3 polymorphism (g.22999655C>A:ECA23), all the animals were homozygous CC, except one Arabian in the high performance group and two Quarter Horses in the low performance index group that were AC. For the PDK4 polymorphism (g.38973231A>G:ECA4), Arabians showed a significantly greater frequency of the G allele than Quarter Horses ($p<0.01$). However, there was no difference between the groups of performance for both breeds. In summary, it was not possible to determine the influence of polymorphisms in the DMRT3 gene in the athletic performance of these breeds since the alleles were almost fixed. Considering the kind of exercise practiced by each breed, it is likely that the PDK4 polymorphism directly influences the control of the pathway to energy production, since endurance Arabians uses predominantly aerobic pathway to energy production and racing Quarter Horses uses the anaerobic pathway.

127

A simple test for genome-wide evidence of selection on complex traits

Dr. Timothy Beissinger[1], Dr. Aaron Lorenz[2], Dr. Jochen Kruppa[3], Professor Henner Simianer[4]
[1]USDA-ARS/University of Missouri, [2]Department of Agronomy and Plant Genetics, University of Minnesota, [3]Institute for Animal Breeding and Genetics, University of Veterinary Medicine, [4]Department of Animal Sciences, Georg-August-University

Recent studies have shown that hard sweeps corresponding to positive selection on new mutations are not the primary driver of evolution for various species including humans and maize. Instead, there is increasing evidence that selection on complex traits, or traits that involve many loci of small effect, is likely to play an important role. To evaluate this prediction, we developed a simple test statistic, applicable to multi-generation experimental or breeding populations, to test for selection on polygenic traits. Our test is based on identifying relationships between allele frequency changes over time and allelic effects, which can be estimated from genome-wide association studies or via linear models with shrinkage. The simple nature of our test statistic allows straight-forward permutation testing to assess significance. Therefore, our approach can be used to either 1) confirm that prior phenotypic selection was effective at shaping the genome, or 2) identify traits that may have been under quantitative selection in the past. We implemented this test in the Wisconsin Quality Synthetic maize population, which has undergone five cycles of recurrent selection for improved silage quality. In this population, we observe evidence of selection for several complex silage-quality traits, including digestibility, starch, and protein. The method also proves to be effective in testing for quantitative selection in livestock data and simulated datasets, where we demonstrate that it can identify evidence of selection even when traditional mapping strategies are incapable of detecting individual significant loci.

128

Artificial selection for odor-guided behavior in Drosophila melanogaster

Ms. Elizabeth Brown[1], Mr. Cody Patterson[1], Dr. Stephanie Rollmann[1]
[1]*University of Cincinnati*

In natural populations, efficient and successful foraging and feeding behaviors are required for survival and reproduction. Such behaviors can be influenced by several sensory modalities, including olfaction. Odors enable flies to assess food quality and distinguish among food sources. To gain insight into how changes in the olfactory system mediate attraction and aversion and in what ways these modifications influence other traits, we have conducted a series of artificial selection experiments in which we independently selected for high/attractive and low/aversive behavioral responses to two aromatic compounds for over thirty generations. We then investigated potential correlations with other olfactory-related traits and found significant differences in food consumption, body weight, and triglyceride levels, at times odor- and sex-specific. We subsequently performed RNAseq to identify changes in gene expression between high and low selected lines. This multifaceted approach will facilitate the identification of candidate genes associated with changes in odor perception and its influence on feeding behavior.

130

High density genome scan for selection signatures in French sheep reveals allelic heterogeneity and introgression at adaptive loci

Christina M. Rochus[1], Flavie Tortereau[2], Florence Plisson-Petit[2], Carole Moreno[2], Gwenola Tosser-Klopp[2], Bertrand Servin[2]

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences; GenPhySE, INRA; AgroParisTech, Université Paris-Saclay , [2]GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT

We present an analysis of French sheep populations aimed at enhancing our understanding of adaptation in this species. We analyzed a large dataset of 27 sheep breeds representative of the main French sheep populations, comprising of 691 animals genotyped for 600K SNP markers. We studied population structure with classical approaches: Principal Component Analysis, cluster-based analysis of admixture, and a population tree model allowing for migration events. We also used these models to position French sheep diversity in the context of other European breeds genotyped in the Sheep HapMap project. To detect signatures of selection we used two methods based on outlying differentiation between populations (a single marker approach and its haplotype-based extension) that accounts for the hierarchical structure between populations. We clearly identified northern and southern groups reflecting the two main origins for French breeds. The northern breeds exhibited more drift and grouped with other northern European breeds of the SheepHapMap dataset while the southern breeds grouped with Spanish and Italian populations. We detected signatures of selection within each geographical group and our results included regions under selection containing known candidate genes for morphology, coat colour, scrapie resistance and horns. Our results also showed evidence for adaptive introgression between groups along with allelic heterogeneity at some adaptive loci. We present one example of both introgression and heterogeneity, a region on chromosome six which has candidate gene LCORL (associated with stature in mammals), and an example of allelic heterogeneity in a region on chromosome 14 which includes MC1R (candidate gene for coat colour in sheep) where we could show that three breeds had been selected on different haplotypes which we confirmed by resequencing MC1R in these three populations.

132

Implications of genetic variability using bulk within progenies derived from F2 plants in common bean

Lais Andrade Pereira[1], Scheila Roberta Guilherme[3], Angela de Fátima Barbosa Abreu[2], Magno Antonio Patto Ramalho[3]

[1]Universidade Federal De Lavras/ Purdue, [2]EMBRAPA Rice and Beans, [3]Universidade Federal de Lavras

The bulk within progenies derived from F2 plants is one of the most used breeding methods for common bean in Brazil. A variation from the original proposal is that, the progenies are evaluated in experiments with replication from F2:3 generation. Very often no selection is applied within the progenies, even at final stages, before releasing the new cultivars for farmers. Therefore, this new cultivar might be a multiline. Thus, our objectives are to quantify the variation between plants within each F2:4 common bean progeny, verify if the variation between progenies is equivalent to the variation within progenies and verify if there is natural selection playing role with the endogamy increase. Seeds from ten, taken at random, F2:4 progenies were divided into two sub-samples. The first sub-sample was stored in a cold chamber, however the second sub-samples was proceeded further to F2:5 and F2:6 progenies. Afterwards the F2:4 stored seeds and its correspondent F2:6 progeny were sown to generate F4:5 and F6:7 progenies. Each F2:4progeny generated ten new F4:5 progenies, likewise each F2:6 generated ten new F6:7 progenies. The 100 F4:5 were evaluated using the simple latice design, in two meters plots, with ten seeds/meter. The same procedure was applied to 100 F6:7 progenies. The grain yield was measured and phenotypic and genetic parameters were estimated using plants/progenies that were derived from each of ten F2:4 and also between the progenies. The heritability estimated within F4:5 and F6:7 progenies were variable, although some of this was similar to between F2:4 estimates. The F2:6 overall mean was higher than the F2:4, indicating that there might be ongoing natural selection while increasing inbreeding. The natural selection acted in order to increases grain yield in plants/progenies.

133

Association study of lactation performance in a dairy sheep population using classic GWAS and genomic models

Hao Li[1], Xiao-Lin Wu[2], Stewart Bauck[2], David L. Thomas[1], Thomas W. Murphy[1], Guilherme J.M. Rosa[1,3]
[1]Department of Animal Sciences, University of Wisconsin - Madison, [2]GeneSeek, A Neogen company, [3]Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison

Lactation performance is one of the most important characteristics of dairy sheep and identification of genes affecting lactation traits is critical to understand and improve its genetics. The objective of the present research was to identify and compare SNPs significantly associated with lactation performance in a crossbred sheep population using classic GWAS and genomic models. The experimental population consisted of 1,740 sheep with multiple-year records of 180-day milk yield, fat yield and protein yield. Of these, 226 had ovine 60K SNP genotypes as well. Estimated SNP effects were obtained on the scale of estimated breeding values for the 226 sheep. The classic GWAS analysis identified between 40 and 149 SNPs with Bonferroni corrected P-values <0.05 for each of the three lactation traits whereas the BayesCπ model showed between 989 and 2,418 SNPs with a 95% or higher posterior inclusion probability of having non-zero association effects. Overall, estimated SNP effects obtained from classic GWAS had higher correlations (0.61-0.66) with the ridge-regression BLUP (rrBLUP) estimates than the estimates obtained with BayesCπ (4.4-4.5%). When considering the top 1% of SNPs, the correlations were raised dramatically to 0.90 - 0.92 between classic GWAS and rrBLUP, and from 0.43 to 0.45 between classic GWAS and BayesCπ. There were between 10% and 17.9% of SNPs overlapping among the top 1,000 SNPs and around 27% of SNPs overlapping among the top 10,000 between classic GWAS and rrBLUP. Finally, SNP variance was computed by 2pq(a^2), where a is the estimated SNP effect from ridge-regression BLUP. Based on the variance of the top 100 SNPs or mean SNP variances on each chromosome, genes on chromosomes 1, 7, 11, and 15 had the greatest impact on these three lactation traits. SNPs significantly associated with lactation performance were selected and those in approximation to positional and functional candidates will be discussed.

134

Developing a framework to leverage soybean uniform regional test data for genomic prediction modeling

Liana M. Nice[1], Peng Zhou[1], Aaron J. Lorenz[1]
[1]*University of Minnesota*

Performed annually since 1939, the USDA coordinated Uniform Soybean Tests are replicated, region-wide field trials of elite public soybean breeding lines. Over the course of their existence, these tests have been an integral part of the selection and release of regionally adapted soybean varieties by public soybean breeders across the US and Canada. Despite the large effort expended to produce this data, relatively little examination of the data has been performed across the many years, locations, and genotypes assayed. To take advantage of this rich phenotypic resource, we are developing a searchable, database to house the available records which currently exist in books, pdfs, and discordant spreadsheets. To enhance the utility of the dataset, environmental and pedigree data will be compiled and linked to the database. Furthermore, breeding lines that were included in the Northern States (maturity groups 00-IV) Uniform Soybean Tests over the past ten years are being genotyped using Genotyping by Sequencing (GBS) technology. This will provide a framework for analyzing many aspects of the performance and genetic composition of elite public soybean varieties, both overtime and across various growing environments. In particular, as we expand the available genotypic dataset, we are developing and testing genomic prediction models that incorporate the robust phenotypic dataset of the Uniform Soybean Tests with the interrelatedness among soybean breeding lines in this historical dataset. We intend for this historical training set to be a community resource that will provide a prediction pipeline for soybean breeders across the northern soybean growing region. Currently, phenotypic records from years 1989-2015 have been processed. Over these 27 years, 7,642 genotypes have been assayed over 199 locations.

135

Associations with flowering time, latitude, and climate in switchgrass

Dr. Paul Grabowski[1], Dr. Joseph Evans[2,3,4], Mr. Guillaume Ramstein[5], Dr. Shawn M. Kaeppler[5,6], Dr. C. Robin Buell[2,3], Dr. Yiwei Jiang[7], Dr. Michael D. Casler[1]
[1]USDA-ARS, [2]Michigan State, [3]DOE Great Lakes Bioenergy Research Center, [4]DuPont Pioneer, [5]University of Wisconsin-Madison, [6]DOE Great Lakes Bioenergy Research Center, [7]Purdue University

Switchgrass is a North American perennial grass and emerging bioenergy feedstock, and increasing biomass yields will improve the economic viability of switchgrass as a bioenergy crop. Flowering time is an important determinant of biomass yields in switchgrass because the the majority of biomass accumulates during vegetative growth, so later flowering increases biomass production. However, many of the highest yielding, late flowering switchgrass varieties originate from southern latitudes and their productivity suffers at northern latitudes due to winter mortality. A better understanding of the genetics underlying flowering time and cold tolerance will aid the development of switchgrass with improved biomass yields, particularly at northern latitudes. Past genetic and genomic efforts in switchgrass have been hindered by its large, polyploid genome, but these challenges are ameliorated by recently developed exome-capture, which targets the genic space and can produced gene-level resolution of patterns of genetic variation. Here, we use exome-capture derived genotypes to detect associations with flowering time in a northern latitude association panel consisting of more than 500 switchgrass genotypes. In addition, we use an expanded diversity panel consisting of genotypes from across the range of switchgrass in the United States to detect associations with latitude and bioclimate variables which may identify genes responsible for adaptation to different photoperiods and climates. These results provide insights into the genetic regulation of important bioenergy traits and will inform efforts to improve biomass yields at northern latitudes.

136

Common SNPs explain a greater proportion of complex trait heritability than previously thought

Doug C Speed[1], Professor David J Balding[2,1]
[1]*University College London,* [2]*University of Melbourne*

The SNP heritability of a trait represents the proportion of phenotypic variation explained by SNPs. Accurate calculation of SNP heritability allows us to better understand the genetic architecture of complex traits, tells us how much heritability remains missing, and allows us to predict the success of future GWAS and of SNP-based prediction models. In order to estimate SNP heritability, it is necessary to make strong prior assumptions about the distribution of heritability across the genome. The two most widely used methods, GCTA and LD-Score Regression, both assume the heritability of a genomic region is proportional to number of SNPs it contains. By contrast, LDAK assumes that heritability is proportional to the amount of genetic variation tagged by the SNPs. By analysing data for over 50 traits, including quantitative and disease phenotypes, we demonstrate empirically that the LDAK assumption better captures the architecture of complex traits than that used by GCTA and LD-Score Regression. Estimates of SNP heritability using the LDAK model tend to be at least 50% higher than those from GCTA and LD-Score Regression, and therefore our results indicate that common SNPs explain a substantially higher proportion of heritability than previously appreciated. For example, for Schizophrenia and Ischaemic Stroke, our estimates of SNP heritability are twice as high as currently estimates. We also find that, on average over many complex traits, the heritability of a SNP varies with the square root of MAF (minor allele fraction), in contrast with an implicit assumption in many analyses that heritability is independent of MAF. Thus, average SNP effect size increases as MAF decreases, but less rapidly than has generally been assumed.

137

Parallel evolution of ethanol tolerance found in four populations of Drosophila melanogaster

Quentin Sprengelmeyer[1], Dr. John Pool[1]
[1]*University of Wisconsin-Madison*

Ethanol tolerance in Drosophila melanogaster has been shown to increase with latitude and linked to different genes, most notably ADH. This study focused on ethanol tolerance of flies from their ancestral range (Zambia) compared to populations found in higher altitudes (Ethiopia, South Africa, and Uganda) and latitude (France). To test for ethanol tolerance we exposed the flies to an 8% ethanol solution and measured the survivability over a 48-hour period. Flies from low altitude Zambia were found to be the least ethanol tolerant, whereas the other three sub-Saharan high altitude populations had different levels of tolerance. The flies from France were extremely tolerant, having nearly 100% survivability. To ascertain candidate genes responsible for higher ethanol tolerance, bulk segregant analysis was performed to detect quantitative trait loci (QTL). Each of the ethanol tolerant populations had mainly different QTL peaks and substantial differences were observed between strains from the same population as well. These data suggest that the loci for ethanol tolerance may not be fixed, or else that epistatic interactions significantly alter mapping results. From this study new insights can be gained in the genetics of adaption; in particular the polygenicity of a trait evolution, its genetic predictability between the different populations, and the frequency of selection on new mutations versus standing variation.

139

Evaluation of efficiency of marker-assisted seedling selection in an apple breeding program

Mrs. Sushan Ru[1], Dr. Craig Hardner[2], Dr. Julia Harshman[3], Dr. Paul Sandefur[1], Dr. Patrick A. Carter[1], Dr. Kate Evans[1], Dr. Dorrie Main[1], Mr. Daniel Edge-Garza[1], Dr. Cameron Peace[1]
[1]Washington State University, [2]University of Queensland, [3]University of California, Davis

Marker-assisted seedling selection (MASS) has great potential to improve efficiency of traditional apple breeding. Despite theoretical evaluation of genetic gain in clonally propagated crops, where genetic gain is defined as the increase in average genetic values through seedling selection, no empirical validation has been reported in apple breeding for the overall efficiency of MASS considering genetic gain, cost, and time simultaneously. Such validation is important as scenarios in breeding activities are much more complex than theoretical models. To facilitate more efficient MASS, this study provides the first empirical evaluation of MASS efficiency based on data collected in an apple seedling population. A multi-trait, multi-locus DNA test, which explained 12%, 16%, and 17% of the additive variance of apple acidity, crispness, and firmness, was used in MASS. Two-stage seedling selection was used, where seedlings were first selected based on marker information and then remaining individuals were selected again based on phenotypic information. This study indicated that while genetic gains achieved from MASS were not significantly different from traditional seedling selection, cost savings from MASS compared to traditional seedling selection ranged from 13 to 3% when DNA testing cost ranged from $0.5 to 1.5 per sample.

140

Signatures of selection in sweet maize suggest a single donor shapes modern elite se1 inbreds

Matthew Murray[1], Dr. William F. Tracy[1]
[1]*University of Wisconsin-Madison*

In the late 1970's a new mutation that confers high eating quality in sweet maize was characterized. This sugaryenhancer1 (se1) gene was recently fine mapped. With this information we are now able to examine the history and integration of this major gene into the sweet maize germplasm. The se1 allele was originally reported in a single inbred line, Il677a (Ames 27120), and it is thought to have penetrated a large breeding pool from this single donor. Here we examine the history of this allele in a modern population, and the impacts on a breeding pool. To explore the source and history of se1 in modern sweet maize we compare 134 publicly available sweet maize lines in the Ames panel, to 50 modern high quality inbred lines from the University of Wisconsin-Madison sweet maize breeding program which were genotyped using GBS, and imputed with FILLIN. Across the genome, scans of allele and haplotype frequency changes were done using Fst, IHS, and rSB. Cross-validated imputation accuracies of $R^2 > 0.90$ suggests the method works well in this population. We identify a single longer than expected haplotype surrounding the se1 locus, which is on average 7.8 MB, that matched the Il677a haplotype, supporting a single introduction theory. Unexpectedly, there are many regions of high Fst and extended Il677a haplotype throughout the genome indicating multiple points of selection. A previous target of selection, sugary1 (su1), also has an extended haplotype with low diversity throughout all of the sweet maize germplasm. We have provided evidence for the utility of GBS and FILLIN in modern sweet maize as well as confirmed the source of the se1 allele in a modern commercial pool of germplasm, and provided an example of the integration of important germplasm into breeding programs.

141

Genetic correlation between gain and carcass traits in purebred and crossbred pigs

Mr. Rodrigo M. Godinho[1,2], Dr. Rob Bergsma[3], Dr. Fabyano F. Silva[1], Ms. Claudia A. Sevillano[2,3], Dr. John W. M. Bastiaansen[2], Dr. Egbert F. Knol[3], Dr. Marcos S. Lopes[3], Dr. Paulo S. Lopes[1], Dr. Simone E. F. Guimarães[1]
[1]Universidade Federal de Viçosa, [2]Wageningen University and Research Centre, [3]Topigs Norsvin Research Center

In pig breeding, selection traditionally takes place based on purebred pig (PB) performance at the nucleus level, while pork production typically makes use of crossbred animals (CB). The success of this selection therefore depends on purebred-crossbred genetic correlation (rpc). The majority of the total production cost of a slaughter pig originates from the growing-finishing phase. There is an increasing trend to give attention to the performance of CB in order to better select PB for CB performance. The current study aimed to estimate rpc for daily gain (DG), back fat thickness (BF) and loin depth (LD), and estimate genetic correlations (rg) between those traits in PB and in three-way CB pigs. Data was obtained on 194,402 PB from 24 nucleus farms and 50,807 CB from three experimental farms. The PB population consisted of five sire and four dam commercial lines, and CB population consisted of three-way crosses between those PB lines. Estimates of rpc for DG, LD and BF were 0.61, 0.75 and 0.82, respectively. Estimates of rg between the traits DG and LD, DG and BF, and LD and BF were, respectively, -0.22, 0.24, and -0.07 in PB and 0.10, -0.05, and -0.07 in CB. The magnitudes of these rpc indicate that genetic progress is being realized in crossbreds at the commercial level from selection on purebreds performance at nucleus level but that including phenotypes recorded in CB on commercial farms will lead to higher genetic progress for these traits in CB. The rg between gain and carcass traits change signs between PB and CB, being favorable in CB, what is relevant for a breeding program aimed on the CB performance. It is worth noting, rg between DG and BF in CB was not significantly different from zero using a T-test ($p<0.05$).

142

Genetic parameters for beef tenderness over various post-mortem aging times determined with a random regression model

Dr. Duc Lu[1], Dr. Ira B. Mandell[2], Dr. Stephen P. Miller[1]
[1]AgResearch Ltd, [2]University of Guelph

Tenderness remains one of the primary beef quality concerns with genomic prediction of genetic differences presenting one method for improvement. Beef tenderness is a complex trait with multiple underlying factors including amount and type of connective tissue as well as response to post-mortem aging. Beef tenderness improves with aging, primarily through the activity of proteolytic enzymes. Utilizing tenderness measured across aging times (7, 14, and 21 d), a random regression model was fit in AsReml software to elucidate different genetic components. Cattle were fed at University of Guelph feedlot facilities and slaughtered at the University of Guelph Meat Science Laboratory or a commercial packing plant in Guelph. Beef tenderness was measured via shear-force on longissimus steaks at the University of Guelph Meat Science Laboratory after wet aging. Animals (2,200) were predominantly crossbred with the predominant breeds being Angus and Simmental. The final model fit first order Legendre polynomials with random effects including an additive genetic and permanent effect (repeated measures). Fixed effects included age of animal and contemporary group. The transformed estimated heritabilities at 7, 14, and 21 d were 0.29, 0.21 and 0.14, respectively. The genetic correlation between the intercept and slope of the regression was -0.73, indicating higher initial shear force is associated with a greater rate of decline with aging. Permanent environment effects as a proportion of total variance were 0.52, 0.50 and 0.40 at the three respective aging times. Compared to a multiple trait model, heritabilities with random regression presented decreasing genetic variance over aging times, although heritability at 7 days was greater with the random regression model. This approach presents an alternative analyses where genetic estimates for underlying components of beef tenderness including initial tenderness as well as response to aging could provide better component traits for further genomic studies.

144

GWAS of tetraploid potato with automated genotype calls

Cari A Schmitz Carley[1], Jiwan P Palta[1], Joseph J Coombs[2], David S Douches[2], Jeffrey B Endelman[1]
[1]*Dept. of Horticulture, Univ. of Wisconsin,* [2]*Dept. of Plant, Soil and Microbial Sciences, Michigan State Univ.*

Accurate tetraploid genotype calling for SNP markers in cultivated potato and other autotetraploids is more difficult than in diploids due to the presence of three heterozygous classes. Current methods rely on either visual inspection of the allele signal ratios (i.e., theta values) or automated calling based on a normal mixture model, which has difficulty identifying rare genotypes. To overcome this limitation, a new automated approach based on hierarchical clustering was implemented in the R software package ClusterCall. ClusterCall uses the expected segregation ratios within biparental populations to "learn" the most likely genotype assignment for each cluster of theta values. Markers with consistent theta boundaries across multiple populations are then retained to predict genotypes in a population of arbitrary composition. To illustrate, three biparental potato populations (n = 160, 162, 191) were genotyped with an Infinium 8303 SNP array and called independently. There were 4711 polymorphic SNP markers for which the genotype calls across the three families were at least 95% identical. This training population was then used to predict genotypes for the SolCAP diversity panel (n = 187). When compared with the visual genotype calls, 99.6% of the markers had at least 95% agreement. Genome-wide association studies (GWAS) were conducted using both the automated and manually-called genotypes for 13 traits. For both marker sets, there was strong statistical support for QTL for tuber shape and tuber eye depth on chromosome 10. Only with the ClusterCall marker set, however, were we able to detect a well-documented QTL (from biparental studies) for vine maturity on chromosome 5. The results demonstrate that ClusterCall generates reliable tetraploid marker genotypes, in a way that readily scales up to larger populations and higher marker density.

145

Meat quality traits genome-wide association study in Nellore cattle

Ana F.B. Magalhaes[1], Rafael Espigolan[1], Roberto Carvalheiro[1], Fernando Baldi[1], Gerardo A. Fernandes Junior[1], Daniel G.M. Gordo[1], Tiago Bresolin[1], Gregorio M.F. Camargo[1], Luis A. L. Chardulo[2], Lucia G. Albuquerque[1]
[1]FCAV - UNESP, [2]FMVZ - UNESP

Tenderness and intramuscular fat are the most important meat attributes, determining consumer's product acceptance. These quality traits affect meat juiciness and chewing. The objective of this study was to identify genomic regions that are associated with meat quality traits in Nellore breed. A total of 1,873 Nellore steers, finished in feedlots and slaughtered at a commercial slaughterhouse, were used. Animals were genotyped using Illumina BovineHD BeadChip of 777kb SNP. After quality control, remained 369,722 SNPs, and 1,630 animals with phenotype and genotype for tenderness, and 1,633 for marbling. The ssGWAS (single-step) method was used for genome-wide association study (GWAS). The model included animal effects as random and fixed effects of contemporary group (year of birth, farm and management group at yearling) and slaughter age and days from slaughter to meat physicochemical analysis as covariates (linear effect). The results are presented as the proportion of variance explained by windows of 150 adjacent SNPs. Only windows explaining at least 0.5% of genetic variance were considered. For marbling, the main genomic regions were located on chromosomes 5, 15, 16 and 25 and explained 3,89% of the additive genetic variance. Annotated genes located in these windows are related to carbohydrate and lipid metabolism in other cattle breeds. For tenderness, the genomic regions were located on chromosomes 5, 7, 10, 14 and 21 and explained 3.80% of the additive genetic variance. Some of the annotated genes found in this work were previously described as affecting tenderness in Bos Taurus. Other genes in these regions were related to growth and muscle development in several cattle breeds. This is the first step in the identification of causal mutations for meat quality traits in Nellore cattle.

Keywords: genetic variance, GWAS, indicine cattle, meat quality
Acknowledgments: This work was supported by Sao Paulo Research Foundation (FAPESP) grant #2009/16118-5

146

Multi-breed genomic prediction using single- and multi-trait models

Mario Calus[1], Mike Goddard[2,3], Iona MacLeod[2], Yvonne Wientjes[1], Phil Bowman[3], Ben Hayes[3,4]
[1]Animal Breeding and Genomics Centre, Wageningen UR, [2]Faculty of Veterinary and Agricultural Science, University of Melbourne, [3]Biosciences Research, Department of Economic Development, Jobs, Transport and Resources, [4]School of Applied Systems Biology, La Trobe University

Genomic prediction is applicable across individuals without recent pedigree links, and therefore potentially also across breeds. Empirical results so far, however, show limited to no benefit of combining multiple breeds in genomic prediction. We tested a multi-task Bayesian Stochastic Search Variable Selection (BSSVS) model that was applied previously by other authors for multi-breed genomic prediction. This model accumulates evidence across breeds to indicate whether a SNP locus is linked to a QTL and thus should receive a large or a small effect. SNP effects, in turn, are estimated within breeds. Other models considered in our study were a single- and multi-trait GBLUP type of model, and a single-trait BSSVS model. The same trait in the two breeds was modelled as a different trait in the multi-trait models. The single-trait model used information from one or both breeds as one trait. The data used included a training data set of 6278 Holstein and 722 Jersey bulls, and 374 Jersey validation bulls. All bulls had daughter based phenotypes for milk, fat and protein yield, and genotypes for 474,773 SNPs. Using the same training data, the BSSVS model consistently outperformed the GBLUP model. Also, the BSSVS model generally yielded significantly higher accuracies for Jerseys when Holstein data was added to the training data, while this was not the case for the GBLUP model. The multi-task BSSVS model was not able to outperform the single-trait BSSVS model that used pooled Holstein and Jersey data for training, suggesting a strong correlation between both breeds for traits analyzed. Estimated genetic correlations between the breeds, being 0.86 for milk, 0.40 for fat, and 0.59 for protein yield, confirmed the strong correlation across both breeds. In conclusion, there seems to be some scope for multi-breed genomic prediction, provided that the model used is able to pinpoint underlying QTL.

147

The genetics of just right: Dose-response surface fits to drought and nitrogen limitation applied together allow mapping of maize loci that exhibit nonlinear responses

Ann Stapleton[1], Megan M. Chang[1], Danielle Allery Nail[1], Toni Kazic[2], Susan J. Simmons[1]
[1]*UNCW,* [2]*University of Missouri*

Crop improvement must accelerate with the increased human population and environmental changes; genotype predictions shortens breeding cycle times. Incorporating climate prediction and adding information about genetic architecture to breeding programs can optimize the use of available resources. We analyzed the genetic architecture of 3D response curves for maize responses to abiotic stress mixtures—specifically, for response to drought and nitrogen deprivation in combination. Plant growth was measured in a Zea mays inter-mated recombinant inbred population exposed to carefully designed combinations of drought and nitrogen. Quadratic dose-response curves were fit globally and by polymorphic allele to determine which markers were associated with different dose-response curve shapes. Loci controlling dose response curve shapes exhibited non-linear effects, with the better-performing allele at mid-range combined stress typically different than the allele that provided best performance in extreme conditions.

To move toward mechanistic modeling of the QTL effects on the abiotic stress sensing and response system in plants, we developed a predictive function that models the sub-network responsible. By tuning the parameters that control the shape and position of the function values, we reproduced the response surfaces we observed experimentally. Moreover, we were able to identify parameter combinations that produced a wide range of more desirable phenotypes, including surfaces that would predict better growth under very stressed conditions. We encourage application of our findings to crop protection mixture experiments or abiotic-biotic mixture experiments in elite germplasm.

148

Investigating causal relationships underlying phenotypic traits of two fruit species of the Sapotaceae family via graphical models

Renan M Pinto[1,2], Bruno D Valente[3], Aderaldo B Gazel Filho[4], Roseli A Leandro[1], Guilherme J M Rosa[2]
[1]Departamento de Ciências Exatas, ESALQ-USP, [2]Department of Animal Sciences, UW-Madison, [3]Department of Dairy Science, UW-Madison, [4]Embrapa Amapá

Sapotaceae belong to the extended order Ericales, a clade of morphologically variable angiosperm families, subdivided into five tribes with 53 genera and approximately 1250 species. Most of them occur in tropical and subtropical regions of Asia and South America, although they have also been introduced in other regions of the world. In this study, we investigated the causal networks underlying phenotypic traits of fruits of two species belonging to the Sapotaceae family: mamey sapote (Pouteria sapota) and star apple (Chrysophyllum cainito). For this purpose, we used a constraint- and a score-based algorithms for structure learning of Bayesian networks. The stability of the selected structure was assessed via Jackknife resampling. The dataset refers to quantitative traits measured on fruits from a collection located in Cabiria 6 Botanical Garden of Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Costa Rica, including: fruit weight, fruit length, fruit diameter, peel thickness, peel weight, seed length, seed diameter, and seed weight. Afterwards, we fitted structural equation models to two selected causal structures for each fruit species to verify if they are reasonable approximations of the data generating processes, and also estimated and tested the intensity of the path coefficients by maximum likelihood and Wald test, respectively. Despite being species from differents genus, remarkable similarities can be observed on the infered causal structures. Knowledge of such networks can be of great value in agriculture production for the development of optimal management strategies and selective breeding.

149

Genomic prediction of early stage single-cross hybrids in maize

Dnyaneshwar Kadam[1], Sarah M Pott[2], Martin O Bohn[2], Alexander M Lipka[2], Aaron J Lorenz[1]
[1]*University of Minnesota, Twin Cities,* [2]*University of Illinois, Urbana-Champaign*

Prediction of single-cross hybrid performance has been a major goal of plant breeders since the beginning of hybrid breeding. Genomic prediction has shown to be a promising approach, but only limited studies have examined the accuracy of predicting single cross performance. Most of the studies rather focused on predicting top cross performance using single tester to determine the inbred parent's worth in hybrid combinations. Moreover, no studies have examined the potential of predicting single crosses made among random progenies derived from a series of biparental families, which resembles the structure of germplasm comprising the initial stages of a hybrid maize breeding pipeline. The main objective of this study was to evaluate the potential of genomic prediction for identifying superior single crosses early in the breeding pipeline and optimize its application. To accomplish these objectives, we designed and analyzed a novel population of single-cross hybrids representing the Iowa Stiff Stalk Synthetic/Non-Stiff Stalk heterotic pattern commonly used in the development of North American commercial maize hybrids. The single cross prediction accuracies estimated using cross-validation ranged from 0.40 to 0.74 for grain yield, 0.68 to 0.91 for plant height and 0.54 to 0.94 for staygreen depending on the number of tested parents of the single crosses. The genomic estimated general and specific combining abilities showed a clear advantage over the use of genomic covariances among single crosses, especially when one or both of the parents of hybrids were untested. The accuracies of genomic hybrid prediction in the novel family as evaluated by leave-one-family-out cross-validations were moderate for grain yield and high for plant height and staygreen. Overall, our results suggest that genomic prediction of the performance of single crosses made using random progenies from the early stages of the breeding pipeline holds great potential to re-design hybrid breeding and increase its efficiency.

150

Genetic parameter estimation and genome wide association study for NDV response in chickens

Kaylee Rowland[1], Huaijun Zhou[2], Rodrigo Gallardo[3], David Bunn[2], Susan Lamont[1]
[1]*Iowa State University, Department of Animal Science,* [2]*University of California Davis, Department of Animal Science,* [3]*University of California Davis, School of Veterinary Medicine*

Exotic Newcastle disease causes up to 80% mortality in chickens in developing countries where the virus is endemic, limiting the production of eggs and meat. Genetic resistance, complementary to vaccination in a disease prevention program, has the potential to be an important tool to reduce the impact of NDV. The experiment was replicated across three hatches of a commercial egg-laying line, Hy-Line Brown. Inoculation with NDV, La Sota strain, occurred on day 21 by an ocular-nasal route. Body weights, blood, and tears were collected throughout the study. Virus load was estimated from viral mRNA level in tears by qRT-PCR; systemic antibody response to NDV in serum was measured by ELISA. Genomic DNA was isolated for genotyping via Affymetrix 600K chicken SNP array.

We hypothesize that many genes regulate NDV resistance in chickens and aim to identify genetic markers associated with NDV resistance (reduced viral load, high antibody titer) so these markers can be applied in genetic selection.

Analyses of viral RNA and antibody levels have confirmed response of challenge groups to the virus and lack of response in control groups.

ASReml estimated heritabilities of 0.36, 0.25, 0.04, 0.06, 0.1756, and 0.0622 for hatch weight, day 31 body weight, 0dpi and 10dpi antibody levels, and 2dpi and 6dpi viral load respectively. Preliminary GWAS data obtained using the GenSel program developed at Iowa State University indicated potential QTL for body weight, viral load, and antibody titer phenotypes. This study will address important issues including correlations between NDV resistance and production traits such as growth.

Support: USAID Feed the Future Innovation Lab for Genomics to Improve Poultry, Hy-Line International, and Hatch project #5357

151

Genomic genetic evaluation in young animals using different methods

Rafael L. Tonussi[1], Rafael M. O. Silva[1], Ana F. B. Magalhães[1], Rafael Espigolan[1], Bianca F. Olivieri[1], Fabieli L. B. Feitosa[1], Elisa Peripolli[1], Ignácio Aguilar[2], Fernando Baldi[1]
[1]*State University of São Paulo,* [2]*Instituto Nacional de Investigación Agropecuaria*

The objective of this study was to investigate the application of BLUP and ssGBLUP young male in different scenarios of uncertain paternity, in a simulated beef cattle population. The phenotypes and genotypes data were simulated using the software QMSim 1.0. Ten replicates were performed for two traits and the selected heritability (0.12 for age at first calving - AFC and 0.34 for weight at 550 days – W550) was based on estimates of real data and phenotypic variance of 1.0. The genomic relationship matrix (G matrix) was considering scenarios with 0, 25, 50, 75 and 100% of multiple sire (MS). The simulated genome had a total length of 2,333 cM, 735,293 markers and 7,000 QTLs and it was assumed that QTLs explained 100% of the genetic variance. The number of markers and QTLs ranged per chromosome, and both were randomly distributed over 29 BTA. All markers were bi-allelic, mimicking SNPs present in the bovine commercial panels. QTL allele effects were sampled from a gamma distribution with a shape parameter equal to 0.4. The use of genomic information in the model (ssGBLUP) provided more accurate prediction (ranging from 0.305 to 0.674) than traditional BLUP (ranging from 0.054 to 0.501). The accuracy of evaluation decreases as much as the proportion of MS increases in both studied traits. Through BLUP and ssGBLUP, the AFC accuracy decreased 86.7 and 38.5% respectively, and 86.6 and 45.2% for W550, respectively, however, when all animals are genotyped the accuracy practically remained constant. These results are quite interesting because there is an interest in evaluating young male that have no phenotypic nor progeny information. Genetic evaluations predictions accuracies increase 20% with the inclusion of genomic relationship matrix (scenario 0% MS), for both studied traits. Nevertheless, with an increased proportion of MS this percentage is even bigger.
Key Words: accuracy, multiple sire, simulation

152

Contribution of social (or indirect) genetic effects to health and disease - a study in laboratory mice.

Dr. Amelie Baud[1], Dr. Megan K Mulligan[2], Francesco Paolo Casale[1], Prof. Robert W Williams[2], Dr. Oliver Stegle[1]

[1]European Bioinformatics Institute, European Molecular Biology Laboratory, [2]Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center

Assessing the impact of the social environment on health and disease is challenging. To quantify the contribution of social effects to more than 100 biomedical traits and genome-wide gene expression measured in laboratory mice, we used a genetic approach: we quantified social genetic effects (SGE), the effects on an individual's phenotype that arise from the genotypes of that individual's social partners.

In an experiment with inbred strains we measured traits relevant to behaviour, growth, and wound healing, as well as gene expression in the prefrontal cortex; in a dataset from outbred (Heterogeneous Stock) mice we studied SGE on behaviour, growth, wound healing, basal immunology, hematology, and blood biochemistry, lung function following exposure to metacholine, glycemic response to glucose injection, and gene expression in the hippocampus. In order to accurately quantify SGE, we derived linear mixed models that account for direct genetic effects (effects on an individual's phenotype that arise from its own genotypes), SGE and correlated residuals due to social and shared environmental factors.
Our results show that genetic variation in cage mates (i.e. SGE) contributes substantially to variation in physiological and molecular traits related to anxiety, the immune function, wound healing, and body weight. SGE explained up to 35% of phenotypic variance, and for several traits their contribution exceeded that of direct genetic effects.
Finally, we considered the impact of SGE on classical narrow sense heritability analyses. We found that ignoring SGE can severely bias estimates of direct genetic effects, which may be an important source of "missing heritability" in studies of complex traits in human populations.

In summary, our study represents the first extensive survey of SGE on biomedical traits. Our results improve our understanding of the architecture of complex traits and set the basis for using SGE to study social effects on health related traits.

153

Genomic regions associated with feed efficiency and component traits of pigs fed lower energy, higher fiber diets after divergent selection for residual feed intake

Emily D. Mauch[1], Nick V.L. Serão[2], Jennifer M. Young[3], John F. Patience[1], Nicholas K. Gabler[1], Jack C.M. Dekkers[1]
[1]Iowa State University , [2]North Carolina State Unversity, [3]North Dakota State University

Feed efficiency is an economically important trait for profitable swine production. Over 10 generations, lines of Yorkshire pigs were divergently selected for residual feed intake (RFI), resulting in lines with increased (Low RFI) and decreased (High RFI) feed efficiency when feeding a standard corn-soybean diet. Considering the increased use of feed by-products to help reduce feed costs in the swine industry, a diet with 18% less net energy and 175% more neutral detergent fiber than fed during selection was formulated to explore the effect of diet on RFI and component traits. Phenotypes and Porcine SNP60 Beadchip genotypes were collected on grow-finish pigs (n= 311) from generations 8, 9, and 10 that were fed this alternative diet. The objective of this study was to identify genomic regions associated with RFI and component traits of pigs fed this low energy, high fiber diet. BayesB in GenSel4 software with π =0.9933 was used to conduct genome wide association studies for 6 traits utilizing 46,467 SNP, after quality control. Four small associations on Sus scrofa chromosomes (SSC) 4 and 8 were identified for average daily gain, all of which have previously been reported, but were not identified by Serão et al. (2016) in these lines when fed the standard diet. Average daily feed intake was associated with 7 genomic regions, with the largest association on SSC6, which has not been previously reported. For feed conversion ratio, 4 associations were identified that cumulatively accounted for 12.5% of genetic variation. Two large QTL on SSC2 were associated with backfat depth, which Serão et al. (2016) also identified. Loin muscle area had 1 association on SSC7 account for 6% of genetic variation. Six small associations identified for RFI cumulatively accounted for 3.8% of genetic variation; none of which overlapped with regions identified by Serão et al. (2016). In conclusion, RFI is a highly polygenic trait yet component traits appear less polygenic. Genomic regions identified for RFI and component traits for the low energy, high fiber diet were different from those identified for the standard diet. Funded by AFRI-NIFA grant #2011-68004-30336.

154

Genomic prediction within and between subpopulations of the USA national maize inbred collection

Joseph Gage[1], Aaron Lorenz[2], Natalia de Leon[1]
[1]*Department of Agronomy, University of Wisconsin-Madison,* [2]*Department of Agronomy and Plant Genetics, University of Minnesota*

Genomic prediction is a valuable tool for use in breeding programs. Optimizing selection of the training set is essential to maximize gains per cycle in a genomic selection context. In this study, we investigate how genotypic and phenotypic diversity in a highly structured maize population can be leveraged to select optimal training populations for genomic prediction. A panel of maize inbreds containing 2632 individuals was clustered into five groups based on ADMIXTURE results, which correspond closely to the known Stiff-stalk, Non-stiff-stalk, Tropical, Sweet, and Popcorn subpopulations. SNP genotype data at 617,649 loci was used to perform genomic prediction of four traits: growing degree days to silk (GDD_DTS), days to anthesis (DTA), plant height (PHT), and ear height (EHT). Training sets comprised of lines selected from exhaustive combinations of one to five subpopulations were used to predict each subpopulation individually for each trait. Prediction accuracies ranged from .60 to .85 for GDD_DTS, from .60 to .87 for DTA, from .23 to .60 for PHT, and from .38 to .75 for EHT, depending on the training population used. We found that prediction accuracies are often nearly as high when predicting between subpopulations as when predicting within subpopulations.

155

Genetics of skeletal evolution in unusually large mice from Gough Island

Michelle D. Parmenter[1], Melissa M. Gray[1], Caley A. Hogan[1], Irene N. Ford[1], Richard J. Cuthbert[2], Peter G. Ryan[3], Karl W. Broman[1], Christopher J. Vinyard[4], Bret A. Payseur[1]
[1]University of Wisconsin-Madison, [2]The Royal Society for the Protection of Birds, The Lodge, [3]University of Cape Town, [4]Northeast Ohio Medical University

Organisms on islands often undergo rapid morphological evolution, providing a platform for understanding mechanisms of phenotypic change. One example is the skeleton, which shows unusual patterns in island populations and forms an integral part of the vertebrate body plan. Although the genetic basis of skeletal variation has been studied extensively in laboratory strains, particularly in the house mouse Mus musculus domesticus, the genetic determinants of skeletal evolution in natural populations remain poorly understood. We used house mice living on the remote Gough Island, the heaviest wild house mice on record, to understand the genetics of rapid skeletal evolution in nature. Compared to a mainland reference strain from the same subspecies (WSB/EiJ), the skeleton of Gough Island mice has enlarged considerably. We used X-ray images to measure 16 skeletal dimensions (at 5, 10, and 16 weeks of age) that are designed to capture major changes in skeletal anatomy. Quantitative trait locus (QTL) mapping in a large F2 intercross between Gough Island mice and WSB/EiJ reveals a total of 198 QTL that control skeletal dimensions. QTL exhibit modest, mostly additive effects and Gough Island alleles are associated with larger skeletal size at most QTL. A large proportion of QTL co-localize across skeletal traits and body weight, an observation that may be explained by a high degree of pleiotropy. Our results provide a rare portrait of the genetic basis of skeletal evolution in an island population, and position the Gough Island mice as a model system for understanding the mechanisms of rapid evolution in nature.

156

The effect of crossing strategy on genomic prediction in maize

Mr. Brett L Burdo[1], Dr. Natalia De Leon[1], Dr. Shawn M Kaeppler[1]
[1]*University of Wisconsin - Madison Department of Agronomy*

Genomic Prediction of maize hybrid performance is becoming an increasingly important component of the maize breeding industry. Selection involves identification of improved inbred lines and identification of inbred combinations that produce the best hybrids. The goal of this study is to evaluate the value of producing new hybrids of candidate lines early in the selection process, as opposed to first identifying improved inbreds using testcrosses and subsequently identifying hybrid combinations utilizing genomic prediction. An Iodent type synthetic from six founder lines (PH207, PHG29, PHG35, PHG50, PHG72, and PHN11) and a Stiff Stalk type synthetic from six founder lines (B73, B84, LH145, NKH8431, PHB47, and PHJ40) were used as the progenitor populations to produce doubled haploid lines for this study. The doubled haploids were genotyped using the Illumina MaizeSNP50 genotyping array. The Iodent DH lines were testcrossed to PHB47, Stiff Stalk DH lines were testcrossed to PHG72, and DH lines were randomly paired between the groups to produce hybrids. Hybrids were evaluated at two locations in WI in 2014 and 2015. The primary trait used in the prediction evaluation was an index of yield and moisture, while 8 other ancillary traits were also evaluated. The genetic variance within the Stiff Stalk DH lines per se was much greater than the variance within the Iodents. While the variance of the randomly paired lines was highest, they also had the lowest mean, so the top ranking hybrids were largely from the Stiff Stalk testcrosses. Predicted hybrids will be evaluated in 2016 and 2017, and should provide more information on utility of random experimental crosses have compared to testcrossing within the context of a prediction-enabled maize breeding program.
Funding: This project is funded by AgReliant Genetics.

157

Evolution of gene expression in giant mice from Gough Island

Mark J Nolte[1], Colin N Dewey[3], Bret A Payseur[2]
[1]*Laboratory of Genetics; supported by the Genomic Sciences Training Program (NHGRI 5T32HG002760), University of Wisconsin-Madison,* [2]*Laboratory of Genetics, University of Wisconsin-Madison,* [3]*The Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison*

Island colonizing organisms often evolve extreme body sizes. Although the generality of the phenomenon suggests common evolutionary mechanisms, the genetic basis of extreme body size evolution on islands remains poorly understood. A striking case of island gigantism is that observed on Gough Island (GI) where house mice evolved to become the largest wild representatives of their species in just 200 years. To illuminate the genetic basis of metabolic processes contributing to the extreme size of GI mice and to nominate candidate genes involved in gigantism evolution we characterized gene expression evolution in three metabolic organs. A total of 100 RNA-sequencing libraries were constructed for GI mice and a wild-derived strain with size representative of mainland mice (WSB) for five conditions: the liver from embryonic, 2-week-, and 4-week-old mice, and the gonadal adipose depot and hypothalamus from 4-week-old mice. We employed the software package RSEM to align reads and quantify transcripts. Using the R package DESeq2, we detected differential expression between GI and WSB mice at thousands of genes in each condition. The adipose exhibited a significant excess of highly expressed genes in large magnitude fold change classes, indicating elevated activity of the pathways constructed by these genes (including metabolically driven cell growth). Monitoring patterns of liver-specific gene expression across three time points enabled grouping of genes with similar temporal expression profiles, shedding light on the evolution of co-regulated genes acting to promote extreme size evolution. In all conditions assayed, we found a significant association between proximal upstream single nucleotide polymorphisms and differential gene expression, suggesting a role for cis-regulatory evolution in GI metabolism. We identified 91 DE genes within +/- 3Mb of QTL for body size recently mapped in crosses between GI and WSB mice. A subset of these genes harbor fixed sequence differences between GI and WSB mice in potential upstream regulatory regions. This study further establishes the GI mouse as a model for understanding the genetic basis of insular body size evolution and provides candidate genes for potential functional genetic assays that explore variable and abnormal growth and metabolism in vertebrates.

158

Prediction of agronomic traits in maize using new source of data

Ms. Yang Xu[1]
[1]*University of California-Riverside*

The genome selection holds a great promise to accelerate plant breeding via early selection before phenotypes are measured, and it offers major advantages over marker-assisted selection for highly polygenic traits. In addition to genomic data, metabolome and transcriptome are increasingly receiving attention as new data sources for phenotype prediction. Several statistical methods have been applied to predict phenotypes. We now have 100k SNPs, 28796 transcripts and 748 metabolites measured from 368 diverse inbred lines of maize. The main objective of this study is to compare the predictive accuracy of six agronomic traits with three different data sources using six representative methods including four linear models (LASSO, PLS, BLUP and Bayes B) and two non-linear models (SVM-RBF and SVM-POLY). The result shows that BLUP is consistently better than other methods, whereas SVM-POLY appears to be the worst method. LASSO has the best performance using metabolomic data but worst performance using genomic and transcriptomic data. Among the three sources of data, genomic data have the highest predictability followed by transcriptome and metabolome. We also combined all three omic data into a single model to perform a combined prediction. However, the combined prediction shows no advantages over the single data source prediction.

159

Incorporating transcriptomics into genomic selection to predict quantitative resistance to Cassava Brown Streak Disease (CBSD)

Mr. Roberto Lozano[1], Dunia Pino del Carpio[1], Marnin Wolfe[1], Jean-Luc Jannink[1]
[1]*Cornell University*

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 genome regions that have some biological relevance for the trait of interest. As a case study we used data from an experiment evaluating the trancriptome 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.

160

Predicting genetic variance in barley using Genotyping-by-Sequencing (GBS) Markers and Population Sequencing (POPSEQ)

Jeffrey L Neyhart[1], Kevin P Smith[1]
[1]*University of Minnesota*

The ability to accurately predict the genetic variance (V_g) that may result from a cross would be invaluable to breeders. Several methods to predict V_g have been proposed recently that incorporate genomic prediction and virtual bi-parental populations. Some of these methods require marker genetic map positions, information not readily available for newer genomics technologies such as genotyping-by-sequencing (GBS). Here we demonstrate the use of population sequencing (POPSEQ) in barley (Hordeum vulgare L.) to assign genetic positions of de novo GBS markers for predicting V_g. A 183-line training set was established and evaluated in six environments for days to heading, plant height, and grain yield. Forty bi-parental populations were created from parents within this set, and using the R package PopVar, we simulated post hoc 34 of these populations. This package uses marker effects estimated by RR-BLUP to predict the mean (μ), V_g, and superior progeny mean (μ_{sp}) of each virtual population. We used the same marker effects and observed GBS genotypes for 1,020 F_3 progeny across the 34 populations to obtain genomic estimated breeding values (GEBVs), from which population predictions of μ, V_g, and μ_{sp} were made. Correlations between the PopVar and GEBV predictions of V_g were moderate (0.19 – 0.39), but mirrored the cross-validation accuracy of each trait (0.43 – 0.89). Correlations for μ_{sp} were high (0.85 – 0.93) and significant ($\alpha = 0.05$). Five of the 34 bi-parental populations were evaluated in three environments for the same traits. Estimates of V_g were generally larger than predicted, while estimates of μ_{sp} were generally less than predicted but were moderately correlated with the predictions. This pilot experiment demonstrates procedures that will be used to empirically validate the accuracy of PopVar using intentionally selected crosses and larger population sizes.

161

Development and application of a bioinformatics pipeline for genotyping-by-sequencing (GBS) of autotetraploid potato.

Schuyler D Smith[1], Dr. Jeff B. Endelman[1]
[1]*University of Wisconsin-Madison*

Genotyping-by-Sequencing (GBS) is being widely used in diploid crops as a cost-effective technology for marker identification across genomes. For autotetraploid crops, such as potato, the cost-effectiveness of GBS is less certain due to the higher read depth (50–60X) needed to differentiate the three heterozygous genotypes, that is much higher than the depth necessary to confidently call a true variant. Our objective was to develop a bioinformatics pipeline for variant and genotype calling in autotetraploids and apply it to a set of 91 elite breeding potato lines. Single-end sequencing using two lanes, required to reach desired read depth, of an Illumina HiSeq2000 produced 4.6 Gb of quality reads aligned to 1.8% of the 674 Mb potato reference genome (version 4.03). Using the Genome-Analysis-Toolkit (GATK), we identified 62K single nucleotide polymorphisms (SNPs) with at least 200 total reads showing a genomic density of one SNP per 42 bps. The number of SNPs declined rapidly as the minimum average depth was raised, from 14K SNPs with at least 20X per sample to only 3300 SNPs with at least 50X. As expected, the number of variants called using both lanes was more than twice the number with only one lane: the gain of variants with minimum average depth of 50X per sample was sixfold. Cross-validation was used to compare different methods for imputing missing marker data, including k-NN, Random Forest, and a polyploid Hidden Markov Model (HMM). We conclude that GBS is a viable marker technology for autotetraploid potato, and could be combined with phenotypic data to estimate the genomic narrow-sense heritability for several quantitative traits, however its cost-effectiveness compared to the potato SNP array warrants further study.

162

Multivariate index improves genomic selection accuracy and gain for disease resistance in maize

David B. Willmot[1], Xin Li[1], James Johnson[1], Brian D. Foss[1], Jose Osorio-Fortea[2]
[1]*AgReliant Genetics,* [2]*Limagrain*

Multivariate Genomic Selection (MVGS) has improved genome-wide marker effect estimates for several species by analyzing multiple traits concurrently. MVGS exploits even weak trait correlations. It provides improved accuracy in a time and cost efficient manner thus increasing genetic gain from selection among untested genotypes. This study explores the accuracy gained by MVGS over single trait univariate genomic selection for disease traits with the objective to predict and cull susceptible lines with minimal error rates. The traits examined in maize were Gray Leaf Spot (GLS), incited by Cercospora zeae-maydis, Northern Corn Leaf Blight (NCLB), incited by Exserohilum turcicum, and Goss's Wilt (Goss), incited by Clavibacter michiganensis pv. nebraskensis.
The experiment was performed on a panel consisting of 421 diverse lines tested in 7 environments per disease each with two replicates arranged in augmented incomplete block designs. Multivariate mixed model analysis was performed using ASReml and, as a result, breeding value estimates were adjusted for the trait correlations between GLS, NCLB and Goss. Heritabilities were high as a result (0.84 to 0.91). In addition, a weighted index was developed to target multiple traits simultaneously thus preserving genetic variability more than tandem selection or independent culling levels. Genomic selection was performed using over 14.5K SNPs with 10-fold cross validation with 5 replications. The prediction accuracies (rMP) were high for multivariate vs. univariate analyses: GLS = 0.91 vs. 0.80, NCLB = 0.79 vs. 0.76, Goss = 0.79 vs. 0.73, Index 0.85 vs. 0.79. Multivariate analysis resulted in higher correlations between the genomic estimated breeding values (GEBVs) and the combined Index than Univariate analysis. This provides better confidence for breeders to make gains for all diseases at once during early stages within a breeding program prior to incurring the cost of field evaluation.

163

The impact of rare alleles on expression and fitness in maize

Karl Kremling[1], Shu-Yun Chen[1], Mei-Hsiu Su[1], Jason Wallace[2], Nicholas Lepak[3], Joshua Budka[3], Peter Bradbury[3], Edward Buckler[3]
[1]Plant Breeding and Genetics - Cornell University, [2]Crop and Soil Sciences - University of Georgia, [3]USDA-ARS - Cornell University

Rare and singleton alleles pose a unique problem in quantitative genetics because their effects cannot be quantified without near infinite sample sizes. However, because selection acts to keep alleles with severe deleterious consequences at low frequencies, these rare alleles are predicted to have an outsize impact on phenotype. Using a burden test which ties rare cis variants to aberrant expression (pioneered by Zhao et al 2016) we show that the abundance of rare alleles in putative regulatory regions leads to extreme expression values which we further tie to significantly lower fitness. This result links the abundance of rare deleterious variants (genetic load) to performance in an artificially selected agronomic species. Thus, despite intensive breeding efforts, rare alleles with severe deleterious effects remain ubiquitous and continue to impact highly selected agricultural species. To profile expression for this study we sampled 7 tissue types in approximately 300 maize lines for which we have whole genome sequence data. Libraries were produced robotically using a low cost ($25/ sample) RNAseq method. The contribution of common regulatory variants to phenotype as determined by a mixed model variance partitioning approach will also be discussed.

164

Identifying deleterious mutations in maize

Dr. Fei Lu[1], Dr. Edward S Buckler[1,2]
[1]*Cornell University,* [2]*USDA-ARS*

Deleterious mutations are constantly reducing fitness of all species by disrupting genes and functional elements across the genome. They contribute to fitness, inbreeding depression, and, likely, heterosis in maize through complementation. Due to natural selection, deleterious mutations are rare in population. Traditional quantitative genetic approaches suffer from limited resolution (QTL mapping) or insufficient allelic replication (GWAS) to allow accurate effect estimates of these rare deleterious mutations. Motivated by identifying deleterious alleles and purging them from population, we sequenced about 300 maize inbred lines at 10X coverage and conducted variants discovery from the data set. Variant effect estimating tools, including Genomic Evolutionary Rate Profiling (GERP) and Sorting Tolerant From Intolerant (SIFT), were used to predict deleterious mutations in maize. We found deleterious mutations have lower allele frequency than neutral ones. Deleterious mutations are enriched in centromeric regions where recombination rate is low, suggesting recombination is a key factor to purge deleterious mutations. In addition, fewer deleterious mutations were found in elite breeding lines, indicating intensive breeding effort has been effective in removing deleterious mutations. These identified deleterious mutations will be incorporated into genomic prediction for multiple traits. It is anticipated that modeling deleterious mutations will enhance current genomic selection approaches and accelerate breeding process of maize and many other crops.

165

Genome-wide association study of seed protein and oil content in soybean nested association mapping population

Mr. Mohammad Wali Salari[1], Katy M Rainey[1]
[1]*Purdue University*

Abstract
The objectives of this study were to determine genotypic differences for soybean protein and oil concentration and to identify Quantitative Trait Loci (QTL) controlling these two traits in SoyNAM mapping population. A total of 5486 genotypes were evaluated in a Modified Augmented Design (MAD) across four locations during the summer of 2012 and 2013. Seed protein and oil contents were estimated using NIR spectroscopy Perten DA7200 diode array instrument. The genotype, location and interaction sources of variation were highly significant for both traits. Locations explained the highest proportion of variation in protein (38.17%) and oil (35.33%) contents, followed by genotypes, which accounted for 33.88% and 29.35% variation in the two traits, respectively. Heritability estimates on a line mean basis for protein and oil concentrations were 0.85, and 0.84, respectively. The phenotypic correlation between these two traits was -0.61, indicating a negative association between the two traits. Genome-wide association study (GWAS) identified 13 QTLs highly associated with seed protein content distributed over 9 different chromosomes and marked by 52 SNPs. Twenty-two out of the 52 SNPs were located within 39.6-40.2 Mbp region of chromosome 9, which is narrower than 5-8Mbp reported previously. Of the 13 seed protein QTLs, six were novel and were located on chromosomes 11, 13, 14, 15, and 18. Additionally, GWAS results also identified 12 QTLs on 8 different chromosomes tagged by 110 SNPs, significantly associated with seed oil content. Six of these QTLs were novel and were situated on chromosomes 2, 11, 15, and 18. The QTLs detected for protein and oil explained 15% and 23% of the phenotypic variations, respectively. The markers closely linked to the novel QTLs could be used for marker-assisted breeding of these two traits.

167

Virgin Genomics: opportunities for emerging sectors - SNP chips optional

Dr. Suzanne J Rowe[1], Mr John C McEwan[1], Mr Ken G Dodds[1], Dr Rudiger Brauning[1], Miss Tracy Van Stijn[1], Miss Hayley Baird[1], Mrs Hannah Henry[1], Mrs Dianne Hyndmann, Professor Nicolas Lopez-Villalobos[2], Dr Shannon M Clarke[1]
[1]Agresearch Limited, [2]Massey University

High throughput genomic technologies are becoming cheaper and sectors with the least infrastructure are reaping the greatest benefits. Techniques such as genotyping by sequencing (GBS) are often immediately applicable, circumventing time and investment associated with developing fixed arrays. Opportunities abound for many sectors such as ecology, conservation genetics, wild populations and emerging livestock industries. Here, we give an example of the developing dairy sheep sector in New Zealand. A milking flock was set up with a collection of 3000 ewes which had no previous recorded pedigree or phenotypes accessible. After milking for a single lactation, the top 1000 ewes (phenotypic selection) and 300 ram progeny were genotyped using both the SheepLD Consortium Array (15K SNP chip) and a restriction enzyme genotyping by sequencing (GBS) approach. This enabled the estimation of inbreeding coefficients, breed composition, and breeding values directly from genomic information circumventing the need for pedigree information. Anticipated higher profitability from greater accuracy of selection, avoidance of inbreeding, and a greater selection response, easily justified genotyping costs. Whilst the SNP chip was convenient and easily facilitated estimation of breed composition by comparison to ovine Hapmap data, GBS is lower cost, more flexible, requires no previous genomic infrastructure investment (e.g. SNP chip development costs) and in some situations is less susceptible to ascertainment bias.
We compare the two genotyping technologies, estimated genomic relationships and discuss the implications of using both to predict performance in an emerging sector.

168

The value of long term multi-year multi-location trial data to evaluate risk in prediction of early cultivar performance

Dr. Vivi N. Arief[1], Dr Ian H. DeLacy[1], Dr Mark J. Dieters[1], Prof Kaye E. Basford[1]
[1]*University of Queensland*

Average genotype performance across environments (i.e. combinations of location and years) is the best estimate of future performance, therefore analysis of data from multi-year multi-location trials aids in selecting genotypes for release to the industry. However, the success or adoption of new cultivars in the market place often depends on its early performance. When genotype-by-environment interactions (GE) are negligible, genotype future performance is the best prediction of early performance. Unfortunately, GE is usually not negligible, thus adding uncertainty to the prediction of genotype performance. This uncertainty could lead to market failure for cultivars if early performance is poorer than their estimated performance. It also increases farmers' risk due to the variability in cultivar performance.
Three components of GE are commonly estimated from multi-year multi-location trials: viz genotype-by-year (GY), genotype-by-location (GL), and genotype-by-year-by-location (GYL) interactions. The repeatable pattern of genotype performance across locations can be successfully managed through a classification of locations that defines target populations of environments for breeding. In contrast, the pattern of genotype performance across years is normally considered to be unpredictable and hence not manageable. In addition, the number of years of data included in the analysis of multi-year trials is usually limited and often not sufficient to capture across year variability. Should sufficient data be available, pattern analysis (i.e. clustering and ordination) can be used to evaluate the historical patterns of GE in multi-environment trials with a view to improving prediction of early performance.
As an example, pattern analysis was used to determine the discrimination among test genotypes provided by locations and years using data from an historical wheat multi-environment trial. The patterns of GY and GL discrimination were investigated by plotting them into the GYL ordination space. The pattern of genotype performance across years and locations can be examined by plotting into the same ordination space. These analyses enable the prediction of genotype performance in locations and/or years where they were not tested. It is argued that understanding the characteristics of genotype performance across years, as well as across locations can aid in the management of prediction risk in early performance due to GE.

169

Reparametrization‐based estimation of genetic parameters in multi‐trait animal model using Integrated Nested Laplace Approximation

Dr. Boby Mathew[1], Dr. Anna Marie Holand[2], Dr Petri Koistinen[3], Prof. Mikko J Sillanpää[4], Prof. Jens Léon[1]
[1]University of Bonn, [2] Norwegian University of Science and Technology, [3]University of Helsinki, [4]University of Oulu

Multi-trait genetic parameter estimation is a relevant topic in animal and plant breeding programs because multi-trait analysis can take into account the genetic correlation between different traits and that significantly improves the accuracy of the genetic parameter estimates. Generally, multi-trait analysis is computationally demanding and requires initial estimates of genetic and residual correlations among the traits, while those are difficult to obtain. Bayesian methods can be applied to estimate genetic parameters in animal models and such methods can, unlike REML (residual maximum likelihood), be helpful in diagnosing identifiability problems associated with models having multiple random factors. Bayesian estimation of genetic parameters is always computationally challenging especially in multivariate framework. Markov Chain Monte Carlo (MCMC) algorithms are commonly used for the Bayesian estimation genetic parameters in animal models. However, non sampling-based Bayesian inference method, Integrated Nested Laplace Approximation (INLA) can be used as a fast and accurate alternative approximation method to MCMC for Bayesian analysis in mixed linear models with multiple random factors.

We propose novel reparametrization-based INLA approach as a fast alternative to MCMC for the Bayesian estimation of genetic parameters in multivariate animal model. Some of the main advantages of our new proposed approach are (1) To avoid difficulties of finding good starting values for analysis which can be a problem for example in Restricted Maximum Likelihood (REML) (2) Bayesian estimation of (co)variance components using INLA is faster to execute than using Markov Chain Monte Carlo (MCMC) especially when realized relationship matrices are dense. We also illustrate the concordance of the INLA results with the traditional methods like MCMC and REML approaches using simulated data sets and real dataset.

170

Divergence and admixture in Chinese and European Sus scrofa

Minhui Chen[1], Guosheng Su[1], Jinluan Fu[2], Qin Zhang[2], Aiguo Wang[2], Mogens S. Lund[1], Bernt Guldbrandtsen[1]

[1]*Aarhus University,* [2]*China Agricultural University*

The introgression from Chinese pigs to European breeds has been a topic of interest in the study of the origin of domestic pigs. However, a genome-wide investigation with a comprehensive coverage of domestic pigs and wild boars has been lacking. We use Admixture and TreeMix on Porcine 60K data from 1,157 domestic pigs and wild boars from across Europe and China. Consistent with previous reports, the results of this study indicate that most European breeds are influenced by introgression from Chinese pigs. Due to the wider spectrum of breeds used here, we are able to detect that the main source of this introgression is pigs from South Chinese breeds, such as Bamaxiang and Dongshan, rather than from Chinese Meishan pigs. Contributions from East and Southwest Chinese breeds are also detectable. There is evidence of multiple distinct introgression events. The extent of introgression varies considerably between breeds. These results characterize the source breeds of the introgression to European breeds and quantify the contribution of Chinese pigs. Thus the hybrid origin of European pig genome is confirmed. We hope these will facilitate further genomic studies of population structure and selection signatures through explicit incorporation of the source of parts of the genome.

171

Nonparametric genome-wide prediction and variable interaction with Bayesian additive regression trees

Dr. Patrik Waldmann[1]
[1]*Swedish University of Agricultural Sciences*

The goal of genome-wide prediction (GWP) is to predict phenotypes based on marker maps, often through single-nucleotide polymorphism (SNP) chips. A large number of methods have been developed for GWP, but most of them are parametric methods that assumes statistical linearity. The Bayesian Additive Regression Trees (BART) method is a nonparametric sum of regression trees approach where the priors provide regularization. A regression tree consists of three components: a tree structure with internal nodes, decision rules and a set of terminal nodes. The regression trees partitions the predictor space into an unknown function which automatically will model nonlinearities within (dominance) and interactions between (epistasis) SNPs.

In evaluation on real data from pig, BART produced smaller prediction error than the LASSO, Bayesian LASSO (BLASSO), genomic BLUP (GBLUP), reproducing kernel Hilbert space (RKHS) regression and Random Forest (RF). Moreover, the SNPs with the highest variable importance measures were investigated for interactions using partial dependence plots. In conclusion, BART can be considered to be an effective and general method for GWP where the regression trees guarantee a very sparse representation of additive and complex non-additive genetic effect. And even though ensemble regression tree methods are black-box approaches, partial dependence plots can provide useful information regarding SNP interactions.

172

Identifying genetic risk factors for Drug-Induced Liver Injury: Detection of susceptible strains and potential vulnerability QTL in the Collaborative Cross mouse population

Dr. Merrie Mosedale[2], Dr. Yunjung Kim[1], Dr. Yuying Xie[1], Dr. William Brock[5], Dr. Sharin Roth[4], Mr. Greg Keele[1], Mr. Robert Corty[1], Dr. Scott Eaddy[3], Dr. Tim Wiltshire[1], Dr. William Valdar[1], Dr. Paul Watkins[2]
[1]Department of Genetics, UNC School of Medicine, [2]Institute for Drug Safety Sciences, UNC School of Pharmacy, [3]Division of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School of Pharmacy, [4]Otsuka Pharmaceutical Development and Commercialization, Inc, [5]Brock Scientific Consulting

Motivation: Tolvaptan, a vasopressin receptor 2 antagonist, is a promising candidate drug for treatment of Autosomal Dominant Polycystic Kidney Disease (ADPKD). However, it has also been associated with drug-induced liver injury (DILI) in a subset of ADPKD patients during clinical trials. This differential vulnerability to drug side-effects was not identified in earlier rodent studies, but those studies used only a single mouse strain and so could not account for the impact of genetic background. Here we use the Collaborative Cross (CC), a genetically diverse set of recombinant inbred mouse strains, to identify strains that are vulnerable to DILI, and to search for QTL that affect this vulnerability.

Methods: A matched placebo-drug design was used on 45 CC strains. Within each strain, eight mice were randomly assigned to treatment groups (N=4 per group per strain), and biomarkers and expression were collected from whole blood and liver tissues at 24h post-dosing. We then examined these data for strain differences, and performed transcriptional analysis, along with haplotype-based QTL and eQTL mapping.

Results: We saw significant drug-induced elevations of a major DILI biomarker, serum alanine aminotransferase (ALT) in 3 of the 45 CC strains, and follow-up QTL mapping identified a potential candidate region on chromosome 14 (58-70Mb). Transcriptional analysis of drug-induced expression changes suggested alterations in immune signaling and oxidative stress. Whole-genome eQTL analysis suggested that whereas expression levels across treatment groups were mainly regulated by local eQTLs, the eQTLs associated with vulnerability were rare and distal, suggesting regulation is more complex and indirect.

Discussion: Using a genetically diverse, replicable mouse population, we identified three sensitive strains and possible susceptibility genes to DILI. Our findings may guide a targeted, hypothesis-based approach to biomarker discovery in a precision medicine approach to manage the risk of hepatotoxicity in tolvaptan-treated human patients.

173

Translating genomics into competitive products for the market

Dr. Sofia da Silva[1], Dr. Milena Ouzunova[1], Dr. Torben Schulz-Streeck[1], Dr. Sofiane Mezmouk[1], Dr. Carsten Knaak[1]
[1]*KWS SAAT SE*

High throughput technologies are shaping the modern world of plant breeding. These technologies convey an entire new view of genotypes that comprise breeding populations. The use of markers allows breeders to view diversity on a level that is independent of phenotype. In addition to that, markers also aid us to bridge the gulf between genetics and performance. To that end, cost efficient genotyping technologies combined with suitable biostatistics approaches can be implemented to assist in the identification of reliable marker-trait associations. At KWS we take advantage of new genomics technologies to accelerate our breeding schemes and to assist with the integration of traits of interest into well performing varieties and breeding germplasm. Genomic selection is one such approach that has been successfully implemented in our breeding schemes. The integration of genomic selection accelerated product development but also opened the door for further reconsiderations in regards to gain from selection.

174

Development of a phenotypic platform for evaluation of kernel processing quality for Frito Lay corn chip production

Mark Holmes[1], Morrie Bryant[2], Peter Coaldrake[2], Dr. Gabe Gusmini[3], Dr. Rex Bernardo[1], Dr. Candice N. Hirsch[1]

[1]Department of Agronomy and Plant Genetics, University of Minnesota, [2]Dupont Pioneer, [3]PepsiCo

The corn chip industry relies on the development of food grade corn (Zea mays L.) hybrids that meet quality specifications necessary for the post-harvest processing chain. Currently, relatively large scale evaluation of hybrids in pilot plants is required to determine which hybrids can be managed through a given processing chain, as the specific traits and parameters necessary for a hybrid to be successful are not well understood. Additionally, there are minimal standardized methods available for evaluating processing traits in hybrids on a small scale. Using a set of 10 hybrids that have been tested for their ability to be managed in the Frito Lay corn chip processing chain, we evaluated multiple bench top phenotyping methods linked to key steps in the processing chain. Each assay was evaluated to determine its effectiveness as a proxy for kernel processing success. Through this study, we seek to develop a benchtop or small scale phenotyping platform that will aid breeding programs in effectively evaluating and developing for food grade corn hybrids that can be successfully managed in a corn chip processing chain.

Note: Gabe Gusmini is an employee of PepsiCo, Inc. The views expressed in this presentation are those of the author and do not necessarily reflect the position or policy of PepsiCo Inc.

175

Genetics of hybrid performance in maize: QTL detection for biomass production in a reciprocal multiparental design

Dr Héloïse Giraud[1], Mr. Cyril Bauland[1], Dr. Alain Charcosset[1], Dr. Laurence Moreau[1]
[1]*INRA, UMR Génétique Quantitative et Evolution – Le Moulon*

Understanding genetic architecture of hybrid performances is of key importance for allogamous species such as maize (Zea mays L.). We developed two multiparental populations corresponding each to one of the main heterotic groups used for maize silage production in Northern Europe (the dent and flint groups). In each group, four founder lines were crossed to produce six connected biparental populations of segregating lines. These lines (821 and 801 for the dent and flint group, respectively), were genotyped for approximately 20k SNPs and were crossed according to an incomplete factorial design to produce 951 dent-flint hybrids, evaluated for silage performances in eight environments. Hybrid genetic variance decomposition showed a predominance of general (GCA) over specific (SCA) combining abilities. SCA explained between 13.8 and 22.6% of the within-population hybrid variance, depending on the trait. QTL detection was carried out for GCA and SCA using different models considering allelic effects transmitted from each founder lines (linkage analysis) or considering directly SNP alleles (linkage disequilibrium mapping) assuming equal or different effects in each group. In total, between 42 and 54 QTLs were detected depending on the model, among which 12 to 31% presented dominance/SCA effect significant at a 5% individual risk level. Only 16 QTL were detected by all three models illustrating their complementary. Most of the QTL (about 80%) were specific to one group, consistent with the long term divergence between the dent and the flint group. These results open interesting prospects for revisiting with markers the concept of reciprocal recurrent selection.

176

Use of individual fitness in analyzing animal lifetime performance

Dr. Asko Maki-Tanila[1], Mr Alper T Kavlak[2], Ms Marja-Liisa Sevon-Aimonen[3]
[1]*University of Helsinki,* [2]*Cukurova University,* [3]*Natural Resources Institute Finland (Luke)*

One measure for the efficiency of animal production is the lifetime reproduction success (LRS) of females. LRS depends on her herd life length and reproduction success. For the wild population context McGraw & Caswell (1996, Am.Nat. 147, 301-310) proposed a single measure integrating survival and age-specific reproduction of the life-history traits which concept they termed as individual fitness (λ_{ind}). It is computed from considering an individual population projection (Leslie) matrix with age-specific fertilities in the first row and age-specific survival probabilities (0 or 1) on the sub-diagonal. The reproduction output is halved to consider the contributions of both females and males to the offspring production. The fitness λ_{ind} is the largest real root of the characteristic equation of the individual transition matrix. Mathematically there are diminishing contributions of reproduction at later ages to λ_{ind}. Typical animal production records are qualitatively appealing because they contain age-specific reproduction data, and λ_{ind} could be used to express the life-time performance of females instead of combining LRS and survival. We used ewe reproduction data from the prolific Finnsheep breed (ProAgria). The number of discrete lambing years of ewes (in total 2610 from 474 herds) born within a single year (1990) covered 1 to 7 lambings with litter size of 1 to 8. The pedigree consisted of 8563 individuals. The statistical animal model considered the effects of herd and age at 1st lambing. The variance components were estimated with Bayesian approach with an uninformative inverse-gamma prior and solved with R program MCMCglmm. The mode (values bound 95%) of posterior distribution of heritability for λ_{ind} and LRS was 0.236 (0.0722 – 0.440) and 0.218 (0.0312 – 0.423), respectively. The genetic correlation between the traits was very high while practically zero between λ_{ind} and the economically important lamb growth rate reflecting artificial selection and planned culling.

177

Genomic prediction in Coffea canephora using Bayesian polygenic modeling

Mr. Luis Felipe Ventorim Ferrão[1], Dr. Romario Gava Ferrão[2], Dr. Maria Amelia Gava Ferrão[2], Dr. Aymbire Fonseca, Dr. Matthew Stephens[3], Dr. Antonio Augusto Franco Garcia[1]
[1]University of São Paulo (ESALQ/USP), Departament of Genetics, [2]Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), [3]The University of Chicago, Departaments of Statistics and Human Genetics

Simulation and empirical results, have been shown that genomic predictions presents sufficient accuracy
to guarantee success in breeding programs. Nevertheless, the effective implementation depends of the ability to connect phenotypic and molecular informations into trustworthy predictive models. Bayesian framework has been used for this end, given the flexibility which the genetic architecture may be modeled and examined. Hence, this study aims the investigation of predictive models, under a Bayesian perspective, that can be applied in plant breeding programs. Real data from experimental populations of Coffea canephora and SNPs identified by GBS were used to elucidate the genotype-phenotype relationship. Bayesian models with different prior distributions were tested, in order to better description the genetic architecture of the four important traits in the coffee scenario. The models tested were: Bayesian version of the RR-BLUP, BayesA, BayesB, BayesC, BayesR and the Bayesian Sparse Linear Mixed Model. The predictive abilities were assessed using a Replicated Training-Testing evaluation, with 30 repetitions, 100.000 MCMC chain and 10.000 burn-in period. Correlation between predicted and observed values and the mean squared prediction error were used to compare the models. All these methods were implemented on the the dscr package (github.com/stephens999/dscr), in order to conduct these examinations within the dynamic statistical comparisons framework. For traits closely resembled with the infinitesimal model, minimal differences were observed across the models. A slight advantage of variable selection methods was observed in some scenarios. Further, differences in performance were observed across the environments evidencing the GxE interaction magnitude. As practical consequence, it is expected that these results might be useful on the selection of premature genotypes, maximizing the gain selection. Under a theoretical view, opens perspective to better understanding of the genetic architecture and the use of a new computational tool to perform predictions and comparisons among statistical methods.

178

Modeling the variance-covariance matrix of genetic and residual effects in a Panicum maximum genome wide selection experiment

Letícia Aparecida de Castro Lara[1], Mateus Figueiredo Santos[2], Liana Jank[2], Lucimara Chiari[2], Antonio Augusto Franco Garcia[1]
[1]*Luiz de Queiroz College of Agriculture / University of São Paulo (ESALQ/USP),* [2]*Embrapa Beef Cattle*

Panicum maximum stands out among the tropical forage species due to its high biomass yield and excellent nutritional quality, providing high animal performance. The improvement of sexual plant population of P. maximum by recurrent selection increases the probability of releasing superior cultivars. In this work, we aimed to model variance-covariance matrices for genetic (G) and residual (R) effects in a spaced plant experiment to obtain adjusted means that will be used in a recurrent genomic selection program. The experiment was conducted in an augmented block design, with 570 sexual plants as treatments and three cultivars as controls in six blocks. Individual plants were evaluated in eight clippings for total green matter (TGM), leaf dry matter (LDM), stem dry matter (SDM), and regrowth capacity (RC). The analyses were performed using the software GenStat16th. The treatments were separated into two groups: regular, where the effects were assumed as random and controls, where the effects were assumed as fixed. Matrices G and R were analyzed considering five different structures: Identity (ID), Diagonal (DIAG), Autoregressive(1) (AR1), Power (P) and Unstructured (UNST). We also used the kinship matrix between individuals, obtained with the R package AGHmatrix, to model their genetic effects. The model selection was performed based on the AIC and BIC criterion. For all traits, the inclusion of the kinship matrix improved the fit of the model. The best structures obtained for matrix G (genetic effects for clippings) were P for TGM and LDM, DIAG for SDM and AR1 for RC. The matrix R, which contains the residual effects for clippings, blocks and mothers, was modeled using DIAG⊗DIAG⊗DIAG for TGM and LDM, and DIAG⊗DIAG⊗ID for SDM and RC. With these structures defined, it was possible to obtain the adjusted means of these traits, which will be used in a forthcoming genomic selection study.

179

Genome prediction for MIR predicted groups of milk fatty acids in Holstein cattle

Dr. Valdecy Cruz[1], Dr. Mehdi Sargolzaei[2], M.Sc. Luiz Brito[1], M.Sc. Guilherme Nascimento[3], M.Sc. Allison Fleming[1], Dr. Filippo Miglior[4], Dr. Flavio Schenkel[1]

[1]Centre for Genetic Improvement of Livestock, University of Guelph, [2]The Semex Alliance,, [3]Sao Paulo State University, [4]Canadian Dairy Network

Genomic prediction using genome-wide single nucleotide polymorphism markers has been shown to be a powerful tool to capture small QTL effects dispersed over the genome for predicting an individual's genetic potential for quantitative traits in livestock. This study was carried out to investigate the possibility of implementing genomic selection for MIR predicted group of milk fatty acids in Holstein cattle. Genotyped animals with Medium-Density (50k) imputed to High-Density (777k) Bovine BeadChip panel were split into a training (N=5,355) and a validation (N=1,338) set depending on their birth year to evaluate the accuracy of genomic prediction. Animals born before October 24, 2011 were assigned to the training set and the animals born on and after October 24, 2011 were considered as the validation set. De-regressed estimated breeding values from a repeatability animal model for 5 groups of milk fatty acids were used for predicting genomic breeding values of the validation animals using GBLUP methodology. The accuracies of genomic prediction were measured as correlations between predicted genomic breeding values and de-regressed estimated breeding values. The accuracies in the validation set for saturated, unsaturated, long including fatty acids with 17 to 22 carbons, medium including fatty acids with 12 to 16 carbons and short chain including fatty acids with 4 to 10 carbons 4 to 10 carbons were 0.52, 0.42, 0.49, 0.53 and 0.55, respectively. These results show that moderately accurate genomic prediction for MIR predicted groups of milk fatty acids is possible in Canadian Holsteins. Further research should look at other dairy cattle breeds.

180

Investigating the genetics of selection response for flowering time in a multi-environment parallel selection experiment

Heather K Manching[1], Michael Dumas[1], Subhajit Sengupta[2], Yuan Ji[2], David Wills[3], Natalia de Leon[4], Sherry Flint-Garcia[3], James Holland[5], Nick Lauter[6], Seth Murray[7], Wenwei Xu[8], Randall J Wisser[1]
[1]Plant and Soil Sciences, University of Delaware, [2]Program of Computational Genomics & Medicine, NorthShore University HealthSystem, [3]USDA-ARS, Division of Plant Sciences, University of Missouri, [4]Department of Agronomy, University of Wisconsin, [5]USDA-ARS, Department of Crop Science, North Carolina State University, [6]USDA-ARS, Department of Plant Pathology and Microbiology, Iowa State University, [7]Department of Soil and Crop Sciences, Texas A&M University, [8]Texas A&M University AgriLife Research and Extension Center at Lubbock

Genetic diversity is fundamental to the success of crop improvement via plant breeding. Maize contains a vast amount of diversity that can be leveraged in breeding; however, the path to using much of this diversity is challenged by heritable maladaptive phenotypes that mask favorable variation. Therefore, elucidating the genetic response to selection and the genetic basis for adaptation traits, such as flowering time, is crucial for maize improvement. We created a TROPICal Synthetic (TROPICS) population of maize from seven tropical inbred lines and subjected it to selection for early flowering time across a latitudinal transect (eight locations spanning from Puerto Rico to Wisconsin) for two generations. Using a standardized protocol, selection was performed that maintained a reasonably large effective population size: 10,000 individuals were evaluated independently at each location in each generation, and 500 of the earliest flowering individuals were selected and pollinated using bulk pollen to minimize family structure. Following selection, remnant seed from all 16 of the TROPICS selected populations and the original founder population were evaluated in all of the eight original selection sites, providing insight into local versus broad adaptation. Moreover, at each location and generation, leaf tissue was collected from the earliest and latest 500 individuals along with a random sample of 1,000 individuals for genotyping. A genotyping-by-sequencing method was developed and vetted for use in heterozygous samples and applied in genotyping the TROPICS materials. Using extreme mapping of ~384 individuals sampled from the tails within each generation (thus far, two locations and two generations) several highly significant associations were detected. We will present the latest results from this study on the genetics of multi-environment, parallel responses to selection.

181

Genome-based prediction of testcross performance with different testers in hybrid rye breeding

Christina Lehermeier[1], Manfred Schönleben[1], Eva Bauer[1], Hans-Jürgen Auinger[1], Hans-Peter Piepho[2], Andres Gordillo[3], Viktor Korzun[3], Malthe Schmidt[3], Peer Wilde[3], Chris-Carolin Schön[1]
[1]Technical University of Munich, [2]University of Hohenheim, [3]KWS LOCHOW GmbH

Hybrid rye breeding is expected to gain substantially from the use of genome-based selection, where lines are selected based on predicted testcross performance using high-density marker profiles. Selection decisions in hybrid rye breeding are mainly based on the general combining ability of test units crossed with several testers. To assess the optimal allocation of resources between number of lines and number of testers, we evaluated the accuracy of prediction with the same and across different testers. We investigated four consecutive years of data from a Polish rye breeding program. In each year, 249 to 406 S_2 inbred lines were phenotyped as testcrosses with different testers in at least four locations in Poland for grain dry matter yield, plant height, and thousand kernel weight. In most cases we observed high genetic and genomic correlations between testcross performance with different testers. Predictive abilities obtained by cross-validation using genome-based best linear unbiased prediction ranged from medium to high for lines crossed to the same tester and dropped only slightly when prediction was performed across testers. Observed predictive abilities across testers were expected from theoretical considerations following from the high genetic correlations between testers and could be forecasted by a deterministic formula. Our analyses will be extended by investigating local genomic correlations between testers along the genome using multivariate models.

182

Simultaneous modeling of phenotype mean and variance: Rationale, results, and software tools

Robert W Corty[1], William Valdar[1]
[1]*Dept. of Genetics, University of North Carolina-Chapel Hill*

Though there is growing recognition that genetic and non-genetic factors can cause heterogeneity not only in phenotype mean but also in phenotype variance, most gene-mapping studies rely on statistical models that assume homogenous variance. Reasons for this are several: To many geneticists, models that account for variance heterogeneity are unfamiliar, the process and rationale for using these models is hazy, and software for applying them to QTL mapping is inaccessible.

Despite these roadblocks, simultaneously modeling the factors that may influence phenotype variance along with the factors that may influence phenotype mean offers several advantages over homogenous variance models: (1) They can improve the power and false-positive rate of tests for mean-controlling QTL by weighting observations in accordance with their residual variance. (2) They allow for identification of variance-controlling QTL, which may influence phenotype variance directly or signal a GxG or GxE interaction. (3) They allow omnibus testing for QTL that control phenotype mean or variance.

We demonstrate these advantages by reanalyzing an existing data resource and showing how joint modeling not only reveals new, previously hidden, QTL but also clarifies the status of previously marginal detections. Importantly, we establish a standard framework for conducting this type of mapping study in F2 intercross populations, including model-fitting tools, a conditional randomization scheme for FWER control, and a set of visualizations for understanding and interpreting results. These tools are available in R package 'vqtl' on CRAN.

183

Genome-wide family prediction

Dr. Patricio Ricardo Munoz[1], Mr. Esteban Fernando Rios[1], Dr. Marcio Fernando Resende Jr.[2], Dr. Matias Kirst[3], Dr. Marcos D Resende[4], Dr. Janeo E Filho[5]
[1]University of Florida, Agronomy Department, [2]RAPiD Genomics LLC, [3]University of Florida, Genetics Institute, School of Forest Resources and Conservation, [4]Federal University of Viçosa, Departament of Statistics, [5]Federal University of Viçosa, Graduate Program in Genetics and Improvement. Departament of Zootecnia

Genomic selection (GS) is used to compute genomic estimated breeding values (GEBV) of individuals. Implementation of GS in minor crops is limited by the high cost of genotyping each individual of the population. Furthermore, in some crops selection is performed at family, instead of the individual level. Here we studied the implementation of genome-wide family prediction (GWFP) in two loblolly pine (Pinus taeda L.) populations: i) a real breeding population composed of 63 families (5-20 individuals per family), phenotyped for four traits and genotyped using an Illumina Infinium assay with 4740 polymorphic SNPs, and ii) a simulated population, with a similar pedigree, 5000 polymorphic loci and two traits (oligogenic and polygenic). In both populations, phenotypic and genotypic data were pooled at the family level in silico. Phenotypes were averaged across all individuals and allele frequencies were computed for each SNP. Additionally, phenotypic data for each trait was divided into three classes: the smallest 10%, the largest 10%, and values between these extreme categories. Four validation populations were created: i) bottom: 10% families showing the smallest phenotypic values, ii) top: 10% families having the largest values, iii) middle: 10% families showing phenotypes between bottom and top, iv) combined: 3% from families in the bottom, 3% from top, and 5% of families from the middle. Marker effects were estimated at the individual (GEBV) and family (GWFP) levels with Bayes-B using BGLR, and validated using 10-fold cross validation. Predictive ability (correlating phenotypes with GEBV and GWFP) was higher for GWFP in both populations, even after standardizing GWFP to be comparable with traditional GS. In addition, prediction was always higher and more accurate for combined validations. Results revealed great potential for using GWFP in breeding programs that select families, such as outbreeding forage species, due to a significant reduction in genotyping and phenotyping cost.

184

Accounting for genetic effects as confounders in propensity score analysis

Vera C. Ferreira[1], Dr. Bruno D. Valente[1], Dr. Guilherme J. M. Rosa[1]
[1]*University of Wisconsin-Madison*

Assigning causal interpretation to associations obtained from observational data is problematic due to confounding. Genetic effects may play a role as confounders when estimating causal relationships among phenotypic traits due to pleiotropic effects. Propensity score (PS) analysis based on matched samples (MS) seeks to eliminate confounding in observational studies using a potential outcome framework. We investigated the applicability of this approach when estimating the causal relationship between two phenotypic traits in the presence of genetic confounders. One thousand populations were simulated with two phenotypic traits: a binary causal variable and a continuous outcome. The true direct effect of the causal variable on the outcome was θ=3 units. The direct genetic correlation between the traits and heritability for both traits were 0.7 and 0.4, respectively. One thousand genetic markers were simulated, in addition to ten causative loci affecting both traits. Additive and non-additive (pairwise epistatic effects) scenarios were considered. Results were compared with those from marginal and multiple linear regression (MLR) approaches, using MSE, bias and variance of each estimator (var). The LASSO technique was implemented to deal with the p>>n issue in the PS (and the MLR). Matched pairs of individuals with similar PS were established to estimate the causal effect. Results for the additive and non-additive scenarios, respectively, for the PS with MS were MSE=33.91, bias=4.95, var=9.37; and MSE=37.70, bias=4.86, var=14.08; for the marginal regression were MSE=120.99, bias=10.19, var=17.07; and MSE=121.01, bias=10.05, var=20.08 and for MLR were MSE=36.52, bias=5.47, var=6.64; and MSE=39.20, bias=5.50, var=8.50. Results suggest PS with MS provides better estimates of the causal effects when compared to a marginal regression in all scenarios. However, advantages were smaller when compared to MLR. In conclusion, it is important to correct for genetic confounding regardless of the method applied; as ignoring them affects the estimates adversely.

185

Rare variants and parent-of-origin effects on whole blood gene expression identified in large family pedigrees

Dr. Ana Viñuela[1], Dr Andrew A Brown[1], Angel Martinez-Perez[2], Nikolaos I Panousis[1], Dr Olivier Delaneau[1], Dr Andrey Ziyatdinov[2], Dr María Sabater-Lleal[3], Dr Andres Hamsten[3], Dr Juan C Souto[2], Dr Alfonso Buil[1], Dr Jose M Soria[2], Dr Emmanouil T Dermitzakis[1]

[1]University of Geneva, Dpt Genetic Medicine and Development, [2]Instituto de Investigaciones Biomedicas Sant Pau, [3]Karolinska Institutet, Dpt Medicine

Studying genetic regulation of gene expression in related individuals provides insights inaccessible when using unrelated individuals, such as heritabilities, rare eQTL commonly observed in the pedigree, imprinting, and parent-of-origin effects. We recruited 935 individuals from 35 pedigrees, with an average of 27 individuals per pedigree and a total of 8654 related pairs. This study (GAIT2) is an effort to understand idiopathic thrombophilia, for which the cohort has been deeply phenotyped. We sequenced the mRNA transcriptome from whole blood for all individuals. We found a median heritability of expression of 0.22, similar to that estimated in the TwinsUK cohort (h^2 = 0.19, EuroBATS). To better understand the genetic regulation of expression, we identified 7,285 eQTL (corrected $p<0.01$, $\pi1 = 0.86$) using a cis association mapping (MAF>0.01), with variance components to consider the familial structure in the data. To test whether the eQTL discovered could be rare in the general population, we looked at the MAF of the variants in the CEU, GBR, IBS and TSI populations of 1000 Genomes. Compared to eQTL from the Depression Genes and Networks study with blood expression from unrelated individuals, we see an excess of variants with MAF < 0.01 (9.6% eQTL compared to 0%, median MAF is 0.11 in GAIT2, 0.26 in DGN). Using the many trios within the pedigrees, we looked for parent-of-origin effects on expression. We performed a cis scan to find variants where the effect of the reference allele in heterozygotes depended on whether it was maternally or paternally inherited (called here parent-of-origin in expression QTL (poeQTL)). We found 2 significant poeQTL ($p<0.05$) for MEG3 and NDN. Both genes are known to be imprinted with maternal and paternal expression, respectively. Additionally, pvalue enrichment analysis showed a $\pi1$ of 0.38, suggesting than parent of origin effects may be more widespread than previously anticipated in humans. In further work exploring parent of origin effects, we intend to use expression levels from trios to identify examples of imprinting, and look at mechanistic aspects of all observed parent of origin effects by integrating chromatin variation data from reference samples.

186

Predicting causal variants affecting expression using whole genome sequence and RNA-seq from multiple human tissues

Dr. Andrew A Brown[1], Dr Olivier Delaneau[1], Professor Emmanouil Dermitzakis
[1]*University of Geneva*

Genome-wide association studies have produced thousands of genetic regions associated with traits, but identifying the particular variant responsible has proved more difficult. Systematically identifying causal variants needs full ascertainment of all SNPs within a region, which is possible with whole genome sequencing but current studies are limited by cost. However, this is feasible when studying gene expression, as relatively small sample sizes are sufficient to provide thousands of associations. We present work which has combined RNA-seq data from 4 tissues and whole genome sequence for 520 individuals. We show that whole genome sequence does not dramatically increase power to find associations relative to genotype data from arrays (26,351 eQTL discovered compared to 27,659 with sequence). However, sequence eQTL were more enriched in all relevant tissue open chromatin regions called by the Roadmap Epigenomics Project (with odds ratios 1.2 to 1.4, P = 2.6e-3 to 1e-10, 30/31 Bonferroni significant). This suggests that with sequence data, the causal variant is more often identified. Building on these results, we developed a method (CaVEMaN), based on non-parametric resampling, to assess the likelihood of a specific variant being causal. We compare this to other methods, such as considering the difference between peak and second association. In simulated data CaVEMaN provides a slightly conservative estimate of the probability that a variant is causal. In addition, CaVEMaN more often predicts whether the peak eQTL is causal than other methods, as it accounts for phenotypic variability and full linkage structure in the region. This second result was confirmed independently using real data; the top 10% of eQTL predicted to be causal by CaVEMaN were enriched in open chromatin regions relative to those ranked by difference in top P values (OR 1.01 – 2.00, 7/31 Bonferroni significant).

187

Genome-wide association study of milk quality traits in Brazilian water buffaloes using phenotypes from non-genotyped animals

Ms. Camila da C. Barros[1], Mr Daniel J. A. Santos[1], Mr Rusbel R. Aspilcueta-Borquis[2], Mr Guilherme Rosa[3], Mr Humberto Tonhati[1]
[1]São Paulo State University, [2]Universidade Federal da Grande Dourados, [3]University of Wisconsin

Genome-wide association studies (GWAS) are an important step in gene mapping efforts. The detection and estimation of quantitative loci contribute to a better understanding of the genetic component underlying complex traits and to the development of more efficient breeding strategies in agriculture. However, due to budget and logistic constraints in livestock studies, generally only a fraction of the animals in a population is genotyped. In such cases an alternative is to use statistical models that combine the phenotypic information of genotyped and non-genotyped individuals, such as the single-step genomic BLUP (ssGBLUP). The objective of this study was to identify chromosomal regions potentially related to milk yield (MY), fat (%F) and protein (%P) percentages in Murrah buffaloes raised in Brazil using the ssGBLUP approach. The number of animals with phenotype was 3,367, the pedigree file contained 15,495 animals, of which 322 were genotyped with specific 90k panel (Buffalo Axiom® Genotyping Array) of Affymetrix. A repeatability animal model was used. The following criteria of SNP exclusion were used: call rate of sample <0.95, call rate of individual <0.90, MAF <0.05 and HW Equilibrium p<10-6. An iterative process was performed to derive SNP weights as function of squares of SNP effects and allele frequencies. The top 10 SNPs capturing the highest proportion of variance explained were further investigated. The proportions of variance explained by the top 10 SNPs for MY, %F and %P were: 7.48, 9.94 and 6.56%, respectively. Such SNPs were located on the chromosomes 6, 7, 16, 17, 20, 21 and 29 for MY; 1, 5, 6, 17, 18 and 26 for %F; and 1, 2, 6, 8, 9, 15, 26 and 27 for %P. Although the most important genomic regions identified in each trait were not exactly the same, some similarities were observed.

188

Genetically and agronomically characterizing a core-set of interspecific breeding lines

Mr. David Eickholt[1], Dr. Thomas E Carter[2]
[1]*North Carolina State University,* [2]*USDA-ARS Soybean Nitrogen Fixation Unit*

Concern often arises within the soybean community over the lack of genetic diversity in the southern U.S. soybean breeding germplasm pool. Despite these concerns, there is an overwhelming amount of freely available genetic material housed within the USDA Germplasm Collection. Of the nearly 22,000 accessions within the germplasm collection, nearly 1,200 belong to the freely-crossing progenitor species, Glycine soja (wild soybean). A significant genetic bottleneck during the domestication of soybean allows for the possibility that useful diversity exists within the progenitor species. Due to the unruly and undesirable agronomic characteristics of the wild soybean, previous efforts to incorporate it into breeding programs have utilized multiple rounds of backcrossing to develop a testable product. Unfortunately, with each round of backcrossing a large percentage of the wild species genome is lost. Since there is no way to effectively model how novel genetic regions will affect soybean performance, breeding efforts should focus on maximizing diversity while maintaining agronomic utility. The objective of this research was to determine if agronomically useful interspecific breeding lines could be developed through a new breeding approach, and to investigate the genetic architecture of said lines. As an alternative to backcrossing, a bulk-breeding, mega-population, pedigree approach was used to develop approximately 225 F4-derived interspecific breeding lines. Breeding lines were yield tested for three years (2008-2010) and then filtered for yield potential (≥ 70% of check variety). Selected breeding lines were then subjected to 6K Infinium SNP genotyping and 16 lines that captured all of the SNP diversity within the wild species were identified. Using the statistical package R, the size and distribution of linkage blocks across these 16 genomes was determined. Results from the genetic analyses will be presented.

189

On the trait-specific correlation matrix of genomic breeding values

Dr. Friedrich Teuscher[1]
[1]*FBN Dummerstorf*

The genomic relationship matrices (VanRaden, 2008) are widely and successfully applied. These relationship matrices are general ones insofar that they are not trait-specific. Trait-specific relationship matrices developed so far have shortcomings since they assume unlinked loci, linkage equilibrium or Hardy-Weinberg equilibrium. Here, a trait-specific genomic relationship matrix is developed, which does not have these strong assumptions. This is done by considering the marker effects as fixed and the matrix of genotype-codes as random. The appropriateness can be shown. For an efficient application further ideas are needed.

190

The influence of organic and conventional production systems on breeding for carrot top height

Charlene Grahn[1], Dr. Erin Silva[2], Dr. Philip Simon[1,3]
[1]University of Wisconsin- Madison Department of Horticulture, [2]University of Wisconsin- Madison Department of Plant Pathology, [3]USDA- Agricultural Research Service, Vegetable Crops Research Unit, University of Wisconsin-Madison

Differences in production methods between organic and conventional systems can influence cultivar traits and subsequently influence breeding goals that are specific to organic agriculture. This study seeks to answer the question: "Is selection for carrot top height in organic production systems effective in a conventional environment, or must selection be made in an organic environment if it is to be relevant to organic growers?" We evaluated four segregating F2 carrot populations of diverse linage and varying top height in replicated trials located in paired organic and conventional fields at the University of Wisconsin- Madison West Madison Agricultural Research Station. Individuals in each population were phenotyped for top dimensions and mass. Significant phenotypic interactions were observed in organic and conventional systems. Biplot analysis was performed to graphically display genotype by environment and genotype by trait interactions and determine the best selection criteria for top size in these populations. To investigate the impact of phenotyping and selection in conventional and organic systems as well as the heritability of the top height trait in the two production systems, individual plants in the six populations with the 5% smallest, intermediate, and 5% largest tops will be genotyped via Genotyping by Sequencing (GBS) and self-pollinated for generation of F3 populations. The results of this research will inform breeders about the heritability of this trait in the two production systems and the need to use organic production systems when developing new cultivars for organic growers.

191

Improving NGS-SNP identification and inference in tetraploids; a white clover example

Dr. Mehul Bhakta[1], Dr. Luis Inostroza[2], Dr. Matias Kirst[3], Dr. Marcio Resende Jr.[4], Dr. Jeffrey Endelman[5], Dr. Patricio Munoz[1]

[1]Agronomy Department, University of Florida, [2]Instituto de Investigaciones Agropecuarias, Quilamapu, [3]University of Florida Genetics Institute, School of Forest Resources and Conservation, University of Florida, [4]RAPiD Genomics, LLC, [5]Department of Horticulture, University of Wisconsin

Molecular breeding studies start by evaluating the correlation between genotype and phenotype, which in turns drives the capability to predict traits of interest based on an individual's genotypic makeup. Next Generation Sequencing (NGS) - based high-density SNP detection has enabled genome wide association study and genomic selection in several diploid crop species. However, in tetraploid species, accurate SNP calling and applications have many challenges including lack of references genome, difficulties in dosage calling, and potential allele dosage effect on phenotype. To address these issues we have developed a pipeline with existing tools for SNP identification and allele dosage calling in tetraploid species without reference genomes using white clover (Trifolioum repens L.) as an example. High quality genotyping using sequence capture was used to identify ~8000 SNP markers in a white clover population of 192 individuals with a maximum of 25% missing data. Simulations indicate that an average of 50x read depth should be used in order to properly call simplex (triplex) from duplex. Results indicated that SNP tetraploid parameterization could provide more accurate genetic structure information compared to when markers were parameterized as diploid. Our preliminary association study using the identified SNPs and structure data detected markers linked to various traits in White Clover only when tetraploid marker parameterization from a high read depth (50x) was used. We will further discuss the effect of read depth in association and genetic structure analyses.

192

Effect of shrinkage on prediction accuracy of genomic selection in tomato population

Phasakorn Fungfoo[1], Dr. Mathilde Causse[2], Dr. Christopher Sauvage[2], Dr. Janejira Duangjit[1]
[1]*Department of Horticulture, Faculty of Agriculture, Kasetsart University, Chatuchak, 10900 Bangkok, Thailand,*
[2]*INRA, UR1052 GAFL, Génétique et Amélioration des Fruits et Légumes, 67 allée des chênes, CS60094, 84143 Montfavet cedex, France*

Genomic selection is a promising marker-assisted selection based application for quantitative traits improvement. Phenotype prediction model in tomato revealed a potential of genomic selection breeding program. In this study, the genomic estimated breeding values (GEBVs) were estimated by the ridge regression best linear unbiased prediction (rrBLUP) statistical model with a training set of tomato representing a broad-based tomato population. The goals are to minimize mean-squared error and to see the ability to use low marker density (100≤m≤3000). Results showed that shrinkage intensity was affected by the number of markers used in the model; and it ranged from 1% to 16% (the lower number of markers, the higher shrinkage intensity). In this tomato population, accuracies predicted by a shrunk approach revealed benefit of shrinkage over a non-shrunk approach in some traits when using a set of low-density markers.

193

Bayesian AMMI analysis of genotype-by-environment interaction in common beans in Paraná- Brazil

Dr. Ines CB Fonseca[2], Dr. Nelson S Fonseca-Júnior[3], Dr. Fernando B Brito[1], Ms. Milena P Euzebio[2], Dr. Leandro SA Gonçalves[2], Dr. Guilherme JM Rosa[1]
[1]University of Wisconsin-Madison, [2]Universidade Estadual de Londrina, [3]Instituto Agronômico do Paraná

The common bean (Phaseolus vulgaris) is a staple food in Brazil and one of the most important sources of protein, especially for people with lower income. Brazil is an important producer, although the national average yield is considered low. One of reasons is the diversity of environmental conditions in which the bean is grown, making Genotype x Environment (GE) interaction an important component to be considered in breeding programs. The Additive Main Effects and Multiplicative Interaction (AMMI) model is a powerful tool for investigating GE effects, and for which statistical inference regarding the bilinear effects can be obtained by Bayesian implementation. This research was carried out to evaluate inbred lines from the IAPAR's breeding program and cultivars of the "Carioca" commercial group in the rainy and dry seasons in Paraná state using a Bayesian AMMI approach. Twenty genotypes of common bean were evaluated in nine environments in each season, between years 2012 and 2014. A Gibbs sampling approach was used to investigate the posterior distribution, with a single chain of 35,000 iterations, burn-in of 5,000 and thinning interval of 2 samples. Overall, the genotypic effects were higher in the rainy season. The genotypes IPR Campos Gerais, LP 08-134, LP 08-186 and LP 08-187 showed the largest genotypic effects in both seasons. IPR Tanguá and LP 08-157 were better in the rainy season, while BRS Requinte was better in the dry season. Higher levels of GE interaction were observed in the 2012/13 crops. Two groups of genotypes and environments were formed in each season. Genotypes that should be recommended considering their stability are Bola Cheia, Carioca, FT 65, Pérola, BRS Notável, BRS Pontal, IAC Alvorada, LP 07-80 and LP 07-157 for the rainy season, and Bola Cheia, IAC Formoso, IPR 139, BRS Estilo and LP 07-134 for the dry season.

194

Improving genomic prediction by adjusting for spatial correlation and genetic competition in cassava field experiments

Dr. Ani A. Elias[1], Prof. Jean-Luc Jannink[1]
[1]*Department of Plant Breeding and Genetics, Cornell University*

Objective
Cassava (Manihot esculenta) root is an important staple food in sub-Saharan Africa. Breeding experiments were conducted in collaboration with International Institute of Tropical Agriculture (IITA) in cassava to select elite parents.

Method
A genomic selection (GS) model can be used to predict genetic value for selecting parents, by taking advantage of the relationship among genotypes through genomic relationship matrix. However, in field experiments, neighboring plots influence each other's yield (root weight) through potential spatial correlation and genetic competition. We characterized the spatial correlation in GS models by distance decay functions based on correlation structures which include autoregressive, exponential, Gaussian, and spherical. To explore competition, a distance decay function affecting the incidence of competition is used while the correlation structure is based on the genomic relationship matrix. Predictability of candidate models was tested through a k-fold cross-validation method. The best model was chosen as the one with the lowest prediction root mean squared error (pRMSE) compared to that of the base model having no correlation and/or competition components.

Result
We found an 18% decrease in pRMSE for dry root weight for the best model where error correlation is explained by Gaussian structure. From simulation studies we found that up to a 22% decrease in RMSE can be achieved with a spatial structure if all the genotypes are replicated at least twice.

Conclusion
Marked improvement was achieved in predictability of GS models by adjusting for the potential influence of neighboring plots in a cassava field experiment.

195

Genome wide association analysis of genetic variants of virus diseases in Cassava

Mr Siraj Ismail Kayondo[2,3], Dr. Pino Del Carpio Dunia[2], Mr. Alfred Adebo Ozimati[1,2], Dr. Yona Baguma[1], Prof. Vernon Edwards Gracen[1,3], Prof Samuel Offei[3], Dr. Robert Kawuki[2], Dr. Jean-Luc Jannink[1,4]
[1]Cornell University, Department of Plant breeding and Genetics, , [2]National Crops Resources Research Institute, [3]West Africa Centre for Crop Improvement, University of Ghana, [4]United States Department of Agriculture, Agricultural Research Service (USDA-ARS)

Cassava (manihot esculenta Crantz), a key carbohydrate source faces unprecedented challenge of viral diseases importantly, cassava brown streak disease (CBSD) and Cassava mosaic disease (CMD). The economic parts of the crop are rendered unmarketable by these viral diseases resulting into mega fiscal setbacks. Genome sequencing of cassava offers several openings for new researchable themes. The purpose of this study is to identify quantitative trait loci (QTLs) that segregate with CBSD and CMD resistance by association mapping. In effect, 830 cassava accessions are being evaluated for CBSD and CMD resistance across 5 seasons in 3 different agro-ecologies in Uganda. These accessions were genotyped with 183,201 SNP markers by genotype-by-sequencing (GBS) to reveal association with CBSD and CMD resistance. Curation of the genomic data at a 5% minor allele frequency (MAF) resulted into 139,996 polymorphic markers for a GWAS analysis. Population structure was accounted for by running a genomic relationship matrix (GRM) performed in R. Several significant hits potentially associated to CBSD and CMD resistance will be identified. The identified SNP markers would be very useful in gene discovery for future marker assisted selection efforts in cassava breeding.

196

Genomic selection of an index trait to develop winter malting barley

Celeste Falcon[1], Kevin P. Smith[2]
[1]*University of Wisconsin,* [2]*University of Minnesota*

Genomic selection (GS) takes advantage of genome-wide marker data to accelerate trait improvement by allowing breeders to select lines in earlier generations and at multiple times per year. For two breeding cycles, we selected lines that were predicted by marker data to have the best values for a selection index that included winter survival, malt extract, grain yield, plant height, and heading date. The selected lines were evaluated for these traits and 10 others that were not under selection in three to six locations in 2014 and 2015. Genome-wide markers used to generate genomic predictions also allowed us to assess genotypic changes across selection cycles. The objectives of this study were to 1) assess rate of gain from GS, 2) evaluate changes in the genetic variance of the population over cycles of selection, and 3) determine the change in genotype frequencies across cycles of selection for markers associated with winter hardiness. While the mean value of the selection index did not improve, winter survival and malt extract significantly increased after two cycles of GS. Although grain yield, plant height, and heading date were not significantly improved, by including them in the index at low weights, we were able to maintain them. Based on principle components analysis, we observed that the genetic variance of the population decreased over cycles of selection and that the population shifted toward increased similarity with the winter parents used to create the population. Markers for several loci known to affect winter hardiness moved toward fixation between the first and second cycles of GS, implying that this breeding strategy can rapidly shift the frequency of large effect genes that have previously been the target of traditional marker assisted selection. This empirical evaluation of GS will inform plant breeders who are considering whether to adopt this breeding strategy.

197

Efficient QTL mapping of additive and dominance effects in outbred multiparent populations

Mr. Gregory R Keele[1], Dr. Jeremy Sabourin[2], Dr. William Valdar[1]
[1]*University of North Carolina-Chapel Hill,* [2]*NIH/NHGRI*

Multiparent populations (MPPs) are model organism populations derived from diverse sets of multiple inbred strains. In many cases, they are intended to act a stable reference points for genetic analysis, and their design and maintenance typically aims to retain both genetic diverse and replicability. To best harness the power of MPPs, however, powerful new statistical methods are needed that can exploit their unique genetic structure. In particular, these populations consist of individuals with genomes that are mosaics of the founder haplotypes. From a mapping perspective this allows detection of QTL based on inferred haplotypes, which has several advantages over SNP association, including the ability to implicitly model local epistasis.

Although modeling this extra information is advantageous, it also results in statistical challenges that must be overcome, particularly within outbred MPPs where the range of possible diplotypes is greatly expanded and the effects of dominance are more likely to be present. In this setting, the assumptions of a pure additive model are hard to justify scientifically but modeling many dominance parameters on typically-sized datasets equally risks overfitting.

To powerfully fit the full model (additive and dominance effects) we propose the DiploLASSO model. Our method models dominance efficiently by parameterizing it as a set of deviations from additivity and to this set applying LASSO-based selection. When no dominance effects are supported by the data, DiploLASSO reduces to the commonly used and accepted additive model. Simulations of outbred populations have shown DiploLASSO to be more powerful at detecting QTL with dominance effects, while maintaining comparable power with the additive model to detect wholly additive QTL. When strong dominance is present, DiploLASSO will better estimate the additive effects, aiding in the design of follow-up experiments for fine-mapping. The DiploLASSO model efficiently identifies and models dominance effects, better accommodating and utilizing multiparent populations.

198

Characterization of genetic diversity, population structure and linkage disequilibrium within and across 35 European maize landraces using high-density genomic data

Manfred Mayer[1], Sandra Unterseer[1], Eva Bauer[1], Natalia de Leon[2], Bernardo Ordás[3], Chris C Schön[1]
[1]Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, [2]Department of Agronomy, University of Wisconsin-Madison, [3]Misión Biológica de Galicia, Spanish National Research Council (CSIC)

Maize landraces can be a useful source of variation for genetic improvement, based on the following considerations: i) At the beginning of the hybrid breeding era, only a small number of landraces served as source for today's elite germplasm, therefore landraces can be expected to contain unexploited allelic variation. ii) Due to their adaptation to specific local environmental conditions landraces are likely to harbor unique favorable alleles for biotic and abiotic stress resistance. iii) Because of their long history of random mating, landraces are expected to have low levels of linkage disequilibrium (LD) and no population substructure, making them ideal for high resolution genetic mapping. However, up to date, these assumptions could not be appropriately tested as most previous studies in maize landraces were based on measurements across landraces with only one or few individuals per accession and/or low-density molecular data.
We created a powerful dataset by genotyping a broad panel of 35 European maize landraces with 609,442 single nucleotide polymorphism markers with each landrace being represented by 22 to 48 individuals. Percentage of polymorphic markers and average nucleotide diversity pi indicated high genetic variation within landraces and Fst values suggested substantial genetic variation between landraces. Population structure analyses supported the assumption of missing substructure within landraces and revealed low levels of admixture between landraces. Genetic clustering of landraces followed their geographical origin. LD estimates will be presented for sampling schemes varying in the number of gametes and populations and will be compared to maize landrace accessions represented in the CIMMYT SeeD-maize GWAS panel [1].

[1] Hearne S. et al. (2014), http://hdl.handle.net/11529/10034

Funding acknowledgement: This study was funded by the Federal Ministry of Education and Research (BMBF, Germany) within the AgroClustEr Synbreed - Synergistic plant and animal breeding (grant 0315528) and by KWS SAAT SE under a PhD fellowship for M. Mayer.

199

Estimation of effective population size from linkage disequilibrium measures in the Spanish Holstein population

M. Saura[1], S. T. Rodríguez-Ramilo[2], Jesús Fernández[1], M. A. Toro[3], B. Villanueva[1]
[1]*Departamento de Mejora Genética Animal, INIA,* [2]*GenPhyse, Université de Toulouse, INRA, INPT, INPT-ENV,*
[3]*Departamento de Producción Agraria, ETS Ingenieros Agrónomos*

Effective population size (N_e) is an important parameter in animal breeding because it determines the rate at which genetic variation is lost. Intense and accurate artificial selection has been practiced over many years in Holstein dairy cattle, resulting in high rates of genetic gain but also in a reduction in N_e and an increase in inbreeding and loss of genetic variation. In this study we have used genomic data to obtain estimates of both ancestral and current N_e from linkage disequilibrium (LD) measures in the Spanish Holstein population. Samples from 10569 Holstein individuals from 13 generations were available. Genotypes from the BovineSNP50K-BeadChip (Illumina) were obtained and a total of 36693 SNPs were used for analysis. Data were divided in different generations and analysed separately. Thus, independent estimates of LD and N_e were obtained for each generation -referred to as temporal replicates. Estimation of N_e was performed from present up to 50 generations in the past. Averaged estimates of N_e across temporal replicates ranged from 295 individuals 50 generations ago to 95 at present. These results are in accordance with estimates of current N_e obtained from the genealogical and molecular rate of inbreeding in previous studies. Our results reveal that the rate of decrease in Ne has not been constant, reaching a plateau (or even slightly increasing) from about 10 generations ago to the present.

200

A statistical framework for leveraging sequence data from supposed technical replicates of an individual to detect experimental errors

Ariel Chan[1]
[1]*Cornell*

Sequencing experiments involve many steps, including sample preparation, library preparation, sequencing and imaging. Errors can occur during any part of the workflow. We propose a statistical framework for leveraging sequence data from supposed technical replicates of an individual to detect experimental errors (e.g., DNA mislabeling and contamination). The proposed method takes (1) estimates of population minor allele frequencies at L biallelic sites and (2) allelic read depths from the k supposed technical replicates of an individual at the L sites as input and presents an estimate of the conditional posterior probability that the k samples originate from {1, 2, …, k} unique source(s), given the observed sequence data and minor allele frequencies, as output. The method requires no genotype calling, imputation, or information regarding marker order making it a suitable tool for studies relying on low- or medium-depth, high-throughput sequencing and for studies involving species with no complete reference genome. We examine the effect of read depth, missing data, the value of L, and minor allele frequencies on algorithmic performance, using both real and simulated data.

201

Corn hybrids stability and adaptability by Bayesian-AMMI statistical procedures for cercosporiose

Dr. Valentina Milani Craft[1], Dr Carlos Alberto Scapim
[1]*Universidade Estadual De Maringá (UEM)*

Gray leaf spot disease, caused by Cercospora zeae-maydis, was identified in 1924 at USA causing many losses in several hybrids, being one of the most destructive diseases of maize. Genotype average performance is due genetic effect under environment effects influence. Thus, genotype x environment (GxE) interaction evaluation is extremely important to recommend the genotype for planting. Data presented in this work refers to severity level of gray leaf spot disease collected in Brazil from 36 maize lines in eight different environments with two replications in each environment. The aim of this study was to evaluate genotypes adaptability and stability across different environments and quantify GxE interaction via Bayesian Additive Main Effects and Multiplicative Interaction models (Bayesian AMMI). AMMI models offer advantages for genotypes selection leading to a more detailed GxE interaction analyses allowing also additive and multiplicative components combination in the same model. Besides, Bayesian inference has been used as an option to solve problems related with variance components estimation for non-normal distribution data. Bayesian inference usage gives more accurate estimates of variance components, genetic parameters, genetic values and genetic gain. Using this approach, we identified the superior genotype that had the lowest level of GxE interaction and with lower disease severity. We also identified locations with higher stability for gray leaf spot disease, less contributing for the observed variability as well as for genotypes presenting specific adaptation.

202

Partitioning variance in maize populations based on genome annotations

Peter J Bradbury[1], Terry M Casstevens[2], Jeffery C Glaubitz[3], Kelly L Swarts[2], Edward S Buckler[1,2]
[1]USDA-ARS, [2]Institute for Genomic Diversity, Cornell University, [3]Biotechnology Resource Center, Cornell University

Variance partitioning using multiple relationship matrices simultaneously and based on classical quantitative genetics approaches can be used to test hypotheses about the importance of sequence variation at different classes of sites. Potentially, sites can be categorized using a variety of genomic annotations. This approach has been popularized by a number of recent papers in the field of human genetics. The analyses presented here use simulation and real data to investigate the ability of this approach to accurately assess the relative importance of different classes of variant sites in existing maize populations with different mating designs. The goal is to be able to use genomic annotations to help identify sites that are likely to contribute to phenotypic variation.

203

Drone-based high-throughput phenotyping for the selection of high-yielding soybean lines

Ms Fabiana F Moreira[1], Alencar Xavier[1], Anthony A Hearst[1], Dr. Keith Cherkauer[1], Dr. Katy M Rainey[1]
[1]*Purdue University*

The annual rate of genetic gains for soybean grain yield is decreasing. A more efficient selection pipeline in the early stages is the proposed way of increasing the rate of genetic gains. The Soybean Breeding program at Purdue University has previously indicated that early-season canopy development in soybean has a strong genetic association to yield and high-throughput drone-based phenotyping of canopy development can predict the yield potential of soybean lines. Usually the selection in early stages of soybean breeding is performed virtually at random, (i.e. without data-driven criteria) due to the extensive number of field plots. This phenotyping bottleneck could be alleviated using drones. However, data collected in a large number of unreplicated entries creates the need to account for spatial heterogeneity in the data. This study aims to select high yield potential soybean lines using yield performance supplemented with drone-based data of canopy development to demonstrate whether new sources of data and analytical approaches can improve selection efficiency in the early generation stages of soybean breeding pipelines. Weekly measurements of canopy coverage from early vegetative to mid reproductive growth and yield were acquired for 4640 progenies rows and 13 checks during the 2015 season. Variance components and breeding values were estimated using the Henderson model fitted with a Bayesian Gibbs sampling algorithm. Kinship matrix was generated using pedigree information and residual correlation matrix using field distance among plots, thus accounting for the heteroscedasticity associated to the field variation. Phenotypic correlation between yield and average canopy coverage (AC) was estimated to be 0.64. Spearman correlation between the breeding values was 0.47. Effectiveness of selection using yield and AC jointly in comparison to yield alone will be evaluated in the 2016 growing season. Current result indicate that stringent selection based on AC in early generations may be a cost-effective way of selection high yielding lines.

204

Applying genomic selection in beef cattle: a deterministic simulation

Dr. Daniel Gordo[1], Dr. Roberto Carvalheiro[1,2], Dr. Gerardo A Fernandes Júnior[1], Dr. Fernando Baldi[1,2], MSc. Rafael Espigolan[1], MSc. Tiago Bresolin[1], Dr. Lucia G Albuquerque[1,2]
[1]Unesp/Jaboticabal, [2]CNPq Researchers

The aim of this study was to compare the effect of applying reproductive technologies combined or not with genomic information on the response to selection in beef cattle, using deterministic models. The current breeding scheme in Nellore cattle in Brazil was adopted as strategy 1 (S1). The genetic gain by year in this traditional scenario was estimated considering that 20% of the top males were used for natural mating (NM). Higher generation intervals, accuracy, and selection intensities were assumed for AI bulls. For females, it was considered that the top 60% heifers were incorporated every year to the herd. In strategy 2 (S2) the fixed time AI (FTAI) was adopted, and the selection intensities of bulls used for AI was higher than in S1. Similar to S2, in strategy 3 (S3) FTAI was also adopted, however, bulls used for NM (top20%) were genotyped. In strategy 4 (S4) the same parameters as in S3 were used and, additionally, bulls used for AI were genotyped. In strategy 5 (S5), we explored using in vitro fertilization with embryos being produced by genotyped donors. In strategy 6 (S6), besides the parameters used in S5, the embryo genotyping was used. All strategies were compared by the genetic gain per year presented as standard deviation units. The use of FTAI in S2 resulted in a small increase of genetic gain compared to S1, 0.16 and 0.13, respectively. By genotyping only the top20% bulls (S3) did not change genetic gain notably (0.17 genetic standard deviation per year). When S4 was used, the rates of genetic gain were 31% and 23% greater than those achieved with S2 and S3, respectively. The genetic gain in S5 was 0.24 standard deviation per year and compared to S4, this scheme resulted in an increase of 15%. Strategy 6 resulted in the highest genetic gain per year (0.27 standard deviations) which was 28% and 12.5% greater than in S4 and S5, respectively. It was possible to obtain higher genetic gains when genomic information was combined to reproductive technologies.
Acknowledgments: São Paulo Research Foundation Grants# 2014/11537-8 and 2009/16118-5.

205

Quantitative genetic analysis of major secreted glycoproteins in a mouse model of asthma

Lauren J. Donoghue[1], Alessandra Livraghi-Butrico[2], Wesley Crouse[1], Joseph P. Thomas[1], Kathryn McFadden[1], William Valdar[1], Wanda O'Neal[2], Richard C. Boucher[2], Samir N. P. Kelada[1,2]
[1]Department of Genetics, University of North Carolina at Chapel Hill, [2]Marsico Lung Institute, University of North Carolina at Chapel Hill

Allergic asthma is a complex disease influenced by both genes and environment. Mucus hypersecretion in the airways is a hallmark feature of asthma, yet the genetic drivers of mucus hypersecretion are almost entirely unknown. We sought to identify genetic regulators of the two major mucin glycoproteins in airway mucus, MUC5AC and MUC5B, using a quantitative genetics approach with the eight founder strains of the Collaborative Cross (CC, n=3-4/strain) and 151 incipient lines of the CC (referred to as "preCC"). We applied a house dust mite mouse model of allergic asthma to these mice and quantified bronchoalveolar lavage expression of mucin proteins. We detected significant variation in MUC5B expression among CC founder strains and estimated the broad sense heritability (H^2) to be ~53%. PreCC lines exhibited a wide range of MUC5B secretion following a normal distribution, consistent with a polygenic architecture for this trait. Quantitative trait loci (QTL) mapping for MUC5B expression in the preCC identified a QTL on chromosome 2 (152-160 Mb, LOD=7.9). We found that MUC5B expression was correlated with the Mus musculus subspecies haplotype from 152-154 Mb with musculus-derived alleles associated with lower expression than domesticus- or castaneous-derived alleles. Subspecies haplotype at this interval explained (h^2) ~27% of the phenotypic variation. We validated the QTL and allele effects at this locus in an independent experiment with fully-derived CC strains. Using lung gene expression data and SNP databases, we identified Bpifb1 as a candidate gene underlying the QTL. We detected a robust increase and co-localization of MUC5B and BPIFB1 in large airway epithelium by immunohistochemistry in response to allergen sensitization and challenge. Gene expression analysis also detected nominally significant increases in Muc5b and Bpifb1 expression. In conclusion, our data indicate that MUC5B is a complex trait and that Bpifb1 may play a novel role in modulating airway MUC5B expression.

206

Investigation of genotype by environment barriers to maize adaptation using near-isogenic lines capturing an allelic series

David M Wills[1], Nicholas Lauter[2], Teclemariam Weldekidan[3], Miriam Lopez[2], Natalia de Leon[4], Sherry A Flint-Garcia[1], James B Holland[5], Seth C Murray[6], Wenwei Xu[7], Randall J Wisser[3]
[1]USDA-ARS Plant Genetics Research Unit, University of Missouri, [2]USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, [3]Department of Plant and Soil Sciences, University of Delaware, [4]Department of Agronomy, University of Wisconsin, [5]USDA-ARS Plant Science Research Unit, North Carolina State University, [6]Department of Soil and Crop Sciences, Texas A&M University, [7]Texas A&M Agrilife Research

Flowering time can act as a strong barrier to the adaptation of a crop to a new environment. For tropical maize, flowering time is typically delayed in long day environments. This photoperiod response may mask the effects of beneficial alleles via pleiotropy, and the selection required for flowering time adaptation may result in a loss of beneficial alleles in coupling phase with photoperiod sensitive haplotypes. To address questions about the nature of pleiotropy, genotype by environment (GxE) interaction, linkage drag and allelic variation in terms of environmental adaptation, we created a population design constructed as a set of Near-Isogenic Lines capturing an Allelic Series (NILAS). A NILAS is a collection of inbred lines differing at a single specific genomic locus for a range of functional alleles. Marker-assisted selection was used to construct NILAS for each of four photoperiod quantitative trait loci (QTL) implicated as primary barriers to maize adaptation to temperate environments. Fourteen distinct tropical x temperate NILAS contrasts were created using seven tropical inbred lines as allele donors and two different temperate inbred lines as genetic backgrounds. Each NILAS captured 12 overlapping introgressions tiled across the QTL region. The NILAS were assessed in eight environments extending from approximately 18° N latitude (Puerto Rico) to approximately 43° N latitude (Wisconsin) in two replicates of a four-way split plot design. We present evidence for an allele series, GxE, phenological pleiotropy, and linked variation based on the evaluation of >20 characteristics. The NILAS is a powerful design to phenotypically, genetically, and ecologically characterize genomic loci. Our findings demonstrate the complexity of quantitative variation within a chromosome based on multi-environment trials and how the NILAS design is extensible for continued studies.

207

Genetic analysis of an intermediate phenotype for recombination rate variation

Richard J Wang[1], Beth L Dumont[1], Bret A Payseur[1]
[1]*University of Wisconsin-Madison*

During meiosis, chromosomes participate in a delicate choreography, exchanging arms and genetic material in the process of recombination. Recombination occurs in all sexually reproducing organisms, is essential for the proper segregation of chromosomes during meiosis, and shuffles alleles onto new genetic backgrounds. Heritable variation for recombination rate exists among individuals, most of which remains unexplained. To provide insights into the genetic mechanisms governing recombination rate variation, we studied the synaptonemal complex (SC), a proteinaceous scaffold that mediates the pairing of homologous chromosomes during recombination. Rates of recombination are strongly correlated with the length of the SC, forming a material link between genetic distance and a physical structure in the cell.

We focused on the house mouse, a model system in which recombination rate and SC length can be quantified in single cells with immunofluorescent microscopy. We developed a high-throughput method to measure SC length and applied it to cytological images of meiocytes from an F2 intercross between strains of Mus musculus musculus and Mus musculus castaneus. With this novel phenotypic data, we mapped three quantitative trait loci (QTL) responsible for variation in SC length – the first known QTL for naturally occurring variation in this trait. Substantial divergence in recombination rate exists between these two subspecies, and previous work (in the same cross) identified loci responsible for this divergence. The QTL with the greatest effect on both recombination rate and SC length lie on the proximal end of the X chromosome. When we mapped QTL for variation in SC length conditioned on recombination rate, we found the locus on the X was no longer significant. This suggests distinct genetic controls underpin these strongly correlated traits. Our results help elucidate the genetic architecture of recombination rate variation and help explain the role that the synaptonemal complex plays in that variation.

209

The inclusion of transcriptome information in genome-wide association studies and genomic selection models to improve the prediction ability

Isabela C. Sant'Anna[1], Rodrigo F. Santos[2], Annette M. Fahrenkrog[2], Jerald D. Noble[2], Marcio M. F. Resende Junior[3], Matias Kirst[4]

[1]Genetics and breeding Graduate Program, Federal University of Vicosa, [2]Plant Molecular and Cellular Biology Graduate, University of Florida, [3]Chief Executive Officer, RAPiD Genomics, LLC, [4]School of Forest Resources and Conservation, University of Florida

Whole-genome association studies (GWAS) have typically been applied to the identification of loci that regulate complex traits such as crop yield and clinical phenotypes. Most of the variants detected have small effects and, collectively, account for a limited fraction of the total genetic variance. The use of transcriptome information in Genome-Wide Association Studies and Genomic Selection models could improve the prediction ability, explain a larger portion of the phenotypic variation and help distinguish the various factors controlling the phenotypic traits. We recently reported the first GWAS conducted in the bioenergy tree crop Populus deltoides. A population of 425 unrelated individuals was genotyped by resequencing 18,153 genes, followed by SNP identification and detection of their association with a suite of bioenergy related traits. A second layer of "omics" data is now being added by transcriptome analysis of differentiating xylem from the same individuals. RNA sequencing (RNA-seq) has been selected for the characterization of the xylem transcriptome of these individuals. Despite its widespread use, challenges remain in RNA-seq data analysis where one of the most important steps is the normalization of the data. There is no consensus for choosing a normalization procedure from among the many existing methods and even small sources of systematic or random error can cause inappropriate results. Here we evaluated several normalization procedures by correlating estimated RNA-seq expression values between biologically replicated samples. The assessment was performed on read counts obtained from the Tuxedo pipeline. The FPKM Cufflinks output was contrasted to the results from TMM, RPKM and Upper quartile approaches of Noiseq, and Quantile. The results so far suggested that there isn't a significant difference among normalization methods for the purpose of identifying loci associated with gene expression.

210

An expanded Wisconsin maize diversity panel and its application in GWAS of flowering time

Mona Mazaheri[1], Brieanne Vaillancourt[2], Joseph Gage[1], Natalia de Leon[1], Manfred Mayer[1], C. Robin Buell[2], Shawn M Kaeppler[1]
[1] *University of Wisconsin,* [2]*Michigan State University*

Genome wide association studies (GWAS) are a powerful method to identify genomic regions associated with important agronomic traits. The objectives of this study were 1) to expand an existing diversity panel of 596 inbreds to enhance gene discovery in association studies; and 2) to test the utility of the expanded population by GWAS of flowering time. A total of 334 inbreds were added to our previous association panel. The entire Wisconsin Diversity population (WiDiv 2.0) consists of 837 inbreds adapted to the upper Midwest region of the United States. The WiDiv 2.0 population was genotyped with 430,947 RNA-Seq based single nucleotide polymorphism (SNP) markers. Admixture analysis assigned the 837 inbred lines to seven subpopulations: stiff stalk (227 inbreds), non-stiff stalk (270 inbreds), Iodent (63 inbreds), sweet corn (35 inbreds), popcorn (15 inbreds), tropical (84 inbreds), and mixed (143 inbreds). GWAS was performed with WiDiv 2.0 using days to pollination data collected in 2014 and 2015. A significant association (P =10^-7.5) was detected within a 50 kb region of a major flowering time candidate gene (zm22). The diversity panel and the genomic data developed in this study will be used to associate genomic variation and phenotypic diversity in our future studies.

211

Exploring properties of genome wide association analyses based on broadly different prior specifications

Chunyu Chen[1], Juan P. Steibel[1], Robert J. Tempelman[1]
[1]*Michigan State University*

Currently popular methods used for genome-wide association (GWA) studies are based on simultaneously fitting all genetic markers in an attempt to account for population structure. A currently popular strategy (EMMAX) infers upon association for the marker of interest by treating its effect as fixed (i.e., flat prior) with all other markers treated as Gaussian random effects. We conjecture that there may be merit to specifying the markers of inferential interest as also sharing the same prior distribution as all other markers in a joint analysis, whether that distribution is Gaussian (rr-BLUP), a heavier-tailed Student t (BayesA) or variable selection specifications based on a mixture of two Gaussian distributions (SSVS). We simulated ten replicated datasets for each of 9 = 3x3 populations defined by three different literature-based QTL distributions (gamma shapes of 0.18, 1.48, or 3) and three different total numbers of QTL (30, 90, or 300) using genotypes derived from 24,661 SNP markers on 928 individuals based on a Duroc-Pietrain F2 cross at Michigan State University. We quantified receiver operating characteristic (ROC) properties for each model by computing the area under the ROC curve (AUC) up until a false positive rate of 0.05 using 10 SNP or 1MB window based inference. We found that SSVS lead to the best ($P<0.0001$) AUC in all cases except for the most oligogenic and skewed QTL effects. EMMAX was generally the most inferior having an average AUC 60% that of SSVS. We also evaluated marginal a posteriori (MAP) approaches to BayesA and SSVS as an alternative to using MCMC; however, these approaches did not appear to be as promising due to their sensitivity to starting values. It appears then that there are considerable advantages to use variable selection specifications in GWA studies.

212

Genome-based analysis of stayability in Nellore cattle using the single-step genomic BLUP approach

Dr. Gerardo A Fernandes Júnior[1], Mrs. Daniela B A Teixeira[1], Dr. Roberto Carvalheiro[1,2], Dr. Daniel G M Gordo[1], Dr. Luciana Takada[1], Mr. Tiago Bresolin[1], Mr. Rafael Espigolan[1], Ms. Danielly B S Silva[1], Dr. Raphael B Costa[1], Dr. Fernando Baldi[1,2], Dr. Lucia G Albuquerque[1,2]
[1]Sao Paulo State University - FCAV/UNESP, [2]CNPq Researcher

Stayability (STAY) is a trait which measures the capacity of a cow to remain calving in a herd until a certain age, given that the cow had the opportunity to reach that age. In beef cattle, STAY is directly related to cow fertility and maternal ability, being one of the most economically relevant traits for the industry. The objective of this research was to estimate genetic parameters and prediction accuracy of breeding values for stayability in Nellore cattle, using genomic information. The database included phenotypic information of 96,811 Nellore cows, born between 1980 and 2009. Genotypic information (463,345 SNPs) of 2950 animals (930 sires and 2020 heifers) was also used in the analysis. Stayability is a binary trait, with 1 for failure and 2 for success. Success means that the cow remained calving in the herd until 65 months of age or beyond. An additive genetic threshold animal model with fixed effects of contemporary group (farm, year and season of birth), and cow precocity (1 for cows with age at first calving equal or below 31 months and 2 for the others), was used to obtain mixed model equations solutions. Two models were considered: 1) applying pedigree-based relationship matrix (A matrix) and 2) using pedigree-genomic relationship matrix (H matrix). Heritability estimates (SD) for STAY were 0.11 (0.01) and 0.14 (0.01) with A and H matrices, respectively. The overall mean of predicted breeding value accuracies was 6.5 % highest using H matrix. The average increase in breeding value accuracies was 5% for females and 9 % for sires. Even with a small proportion of genotyped animals, compared to standard pedigree-based predictions, combining genomic and pedigree information increased genetic variability estimates and prediction accuracies for stayability trait in Nellore cattle. Acknowledgments: São Paulo Research Foundation Grants # 2009/16118-5 and 2015/06140-4.

213

Developing a method to screen for salt tolerance in carrot

Mr. Adam Bolton[1], Dr. Philipp Simon[1,2]
[1]*Department of Horticulture, University of Wisconsin-Madison* , [2]*USDA-ARS*

Carrot (Daucus carota, L.) is a widely grown and economically important vegetable worldwide. Carrot production is restricted by salt-affected soil in many parts of the world where it provides a dietary source of vitamin A. There is evidence that carrots are especially salt sensitive relative to other crops, but little research has been devoted to the quantification and breeding of salt tolerance in carrot. A promising source of genetic diversity for salt tolerance is carrot wild relatives, many of which grow natively in areas with high levels of salt and, according to preliminary data, vary for levels of salt tolerance. Carrot wild relatives are often weeds with small, white, taproots, and provide little direct economic benefit, but do readily intercross with cultivated carrot, allowing introgression of desirable traits into cultivated varieties. However, salt tolerance is a quantitative trait in many species and accurate field screening is challenging. The objectives of this experiment are: (1) to design a controlled system to screen for salt tolerance while minimizing other environmental influences such as moisture and temperature, (2) to compare the efficacy of commonly used physiological measures of salt tolerance in other species (e.g. Na exclusion, Na/K ratio, and leaf chlorophyll content) to biomass reduction in carrot wild relatives, and (3) to evaluate correlations among controlled measurements and with concurrent field experiments in Bangladesh and Pakistan.

214

Maintaining the accuracy of genomewide predictions when selection has occurred in the training population

Sofia P. Brandariz Zerboni[1], Rex Bernardo[1]
[1]*University of Minnesota*

Routine genomewide selection in maize (Zea mays L.) will lead to only a subset of the lines in an A/B biparental cross being eventually phenotyped. If the A/B cross is used as part of the training population for predicting the performance of lines in a future cross with either A or B as parents, the training population would be a selected rather than a random subset of lines. Our objectives were to determine if selection in the training population (1) reduces the response to selection and accuracy of genomewide selection in the test population, and (2) increases the genetic similarity of the 10% best lines in the test population. A total of 969 biparental maize populations were evaluated at 2-15 locations in 2000-2008 for grain yield, moisture, and test weight. The parents of the 969 populations were genotyped with 2911 single nucleotide polymorphism (SNP) markers, and marker data were imputed from lower-density screening of the progeny in each biparental cross. Having phenotypic information on only a selected fraction of the lines significantly reduced the response to selection and prediction accuracy. However, augmenting the training set with the five poorest lines nearly restored the response to selection and prediction accuracy to their original levels. Prior selection in the training population did not increase the genetic similarity (calculated from non-imputed SNP data) of the best lines in the test population. We concluded that including a small number of the poorest lines in a training population is a practical way to maintain the effectiveness of genomewide selection.

215

Genomic selection shows promise for improving winter wheat: Insights from the University of Nebraska-Lincoln wheat breeding program

Dr. Vikas Belamkar[1], Dr. Mary J. Guttieri[4], Dr. Ibrahim El-basyoni[1], Mr. Waseem Hussain[1], Dr. Jesse Poland[2], Dr. Diego Jarquín[1], Dr. Aaron J. Lorenz[3], Dr. P. Stephen Baenziger[1]
[1]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, [2]Wheat Genetics Resource Center, Department of Plant Pathology, Kansas State University, [3]Department of Agronomy and Plant Genetics, University of Minnesota, [4]USDA, ARS, CGAHR, HWWGRU

Genomic selection (GS) can increase genetic gain. The objective of this study is to inspect whether GS can (1) Predict performance of new lines in a trial; (2) Improve accuracy of selection decisions; (3) Recycle elites line earlier to the crossing block; (4) Reduce costs by phenotyping a subset of lines; and (5) Predict performance of lines across locations. This study comprised of 1,100 entries from four independent F3:6 nurseries evaluated each year from 2012-2015 across 8 to 10 locations in Nebraska. Mixed models incorporating spatial variations provided better fit to the grain yield data. Genotype-by-sequencing, quality control and imputations provided 26,925 SNP markers. For each of the years, genomic prediction ability (PAB) was estimated by randomly marking entries as missing in steps of 10% - from 10% to 90% of dataset. Prediction ability (PAC) was also estimated by marking 100% of the entries missing in a year. Average PAB calculated using 10-fold cross validation ranged from 0.229 to 0.552, and PAC varied from 0.167 to 0.282. Further, we tracked entries from each of these four nurseries that were advanced. It was remarkably apparent that lines with “above average genomic estimate breeding value (GEBV) and observed phenotypic value (OPV)” were being retained for longer times in the breeding program. This suggests using GEBV and OPV can improve accuracy of selection decisions and recycle elite lines earlier to the crossing block. Prediction ability estimated with 50% of the entries missing in each year, found more winners (entries) with above average GEBV and OPV. Hence, evaluation of only 50% of the entries in a year to make accurate selections seems possible. Improving PAC from the current value of ~0.20 to >0.37 will trigger examining skipping of a field trial year. GS across locations and for quality traits is in progress.

216

Prospects of genomic prediction for Goss's bacterial wilt and leaf blight resistance using bi-parental populations and a diversity panel of Maize

Amritpal Singh[1], Aaron Lorenz[1]
[1]*University of Minnesota*

Goss's wilt and leaf blight is a bacterial disease of maize (Zea mays L.) caused by gram positive bacterium Clavibacter michiganensis subsp. nebraskensis (Cmn). Goss's wilt has re-emerged as an important disease in the western United States and is spreading to other areas of North America. One possible reason for re-appearance of Goss's wilt could be the increase in susceptibility of germplasm used in commercial maize breeding. Therefore, it is important to identify the sources of resistance to Goss's wilt that can be used to deploy resistance to Goss's wilt into commercial maize hybrids. Genomic prediction could be used to predict resistance to Goss's wilt. The objectives of study were to (i) assess the ability of genomic prediction to predict the resistance of maize inbred lines in a diverse germplasm pool that can potentially be used as sources of resistance, and (ii) assess the prospects of genomic prediction in maize bi-parental populations. Two types of datasets consisting of a diverse panel of 555 maize inbred lines, and three bi-parental populations B73 x Oh43 (189 lines), B73 x HP301 (140 lines), and B73 x P39 (121 lines), phenotyped for Goss's wilt and genotyped with high-density genotyping-by-sequencing (GBS) SNP data were used. Genomic best linear unbiased prediction (GBLUP) model was used in each dataset to assess the genomic prediction accuracy of prediction of Goss's wilt resistance through a ten-fold cross validation procedure. Prediction accuracy of 0.54 was achieved in the diversity panel. Among bi-parental populations, a maximum prediction accuracy of 0.68 was obtained in B73 x Oh43 population. Results from this study indicate that genomic prediction may be helpful to predict the resistance to Goss's wilt in maize.

217

Construction of high density linkage maps and detection of downy mildew resistance: Locus in a Vitis aestivalis-derived 'Norton' population

Mr. Surya Sapkota[1], Mrs. Li-Ling Chen[2], Dr. Shanshan Yang[3], Dr. Lance Cadle-Davidson[4], Dr. Chin-Feng Hwang[2]

[1]University of Missouri & Missouri State University, [2]State Fruit Experiment Station at Mountain Grove Campus, Darr School of Agriculture, Missouri State University, [3]Arizona State University, [4]Grape Genetics Research Unit USDA, ARS

Grapevine downy mildew is one of the most widespread and destructive diseases of grapevine, particularly in viticultural areas with warm and wet conditions. This disease is caused by the oomycete Plasmopara viticola, which damages green tissues and defoliates vines. Traditional Vitis vinifera wine grape cultivars are susceptible to downy mildew whereas several North American and a few Asian cultivars possess various levels of resistance to this disease. To identify genetic determinants of downy mildew resistance in V. aestivalis-derived 'Norton', a mapping population was developed in 2005 from a cross between 'Norton' and V. vinifera 'Cabernet Sauvignon' at the Missouri State Fruit Experiment Station, resulting in 95 hybrid progenies. This population was further expanded to 182 individuals by repeating the same crosses in 2011. A haploid Norton genetic map was constructed with 379 simple sequence repeat (SSR) markers clustered in 19 linkage groups. In collaboration with VitisGen (www.vitisgen.org), a high density linkage map of Norton and the consensus map was constructed with an additional 1,591 single nucleotide polymorphism (SNP) markers generated by genotyping-by-sequencing (GBS). Disease progression and resistance reaction in response to P. viticola was studied 8 days post-inoculation in the given population for two years. A quantitative trait loci (QTL) analysis indicated a resistance locus on chromosome 18 explaining 40% of the total phenotypic variation. Flanking markers closely linked with the trait can be used for marker-assisted selection in the development of new cultivars with resistance to downy mildew.

218

Genome-wide association analyses based on alternative strategies for modeling feed efficiency

Ms. Yongfang Lu[1], Dr. Mike J. Vandehaar[1], Dr. Diane M. Spurlock[2], Dr. Kent A. Weigel[3], Dr. Louis E. Armentano[3], Dr. Charles R. Staples[4], Dr. Erin E. Connor[5], Dr. Zhiquan Wang[6], Dr. Mike Coffey[7], Dr. Roel Veerkamp[8], Dr. Yvette Haas[8], Dr. Mark D. Hanigan[9], Dr. Robert J. Tempelman[1]
[1]Michigan State University, [2]Iowa State University, [3]University of Wisconsin, [4]University of Florida, [5]Animal Genomics and Improvement Laboratory, [6]University of Alberta, [7]Scottish Agricultural College , [8]Wageningen UR Livestock Research, [9]Virginia Tech

Genetic selection for feed efficiency (FE) in dairy cattle is important for economic and environmental sustainability. Although whole genome prediction analyses of emerging consortia datasets enables genomic selection for FE jointly with other economically important traits, it will be ultimately useful to identify important genomic regions and, subsequently, candidate genes for FE. These inferences may depend upon differences in statistical modeling of FE, for example, using a classical residual feed intake (RFI) analysis versus that based on a recently proposed reparameterization of a multiple trait (MT) model analysis of component traits of FE. An international dataset consisting of dry matter intakes (DMI), milk yields and components used to determine milk energy (MILKE), and metabolic body weights (MBW) on 6,937 cows (2,021 non-genotyped and 4,916 genotyped based on 57,347 SNP markers) from 4 different countries (Canada, the Netherlands, the United Kingdom, and the United States) was analyzed under both models (RFI vs MT), to determine differences in inferences between the two models for genome-wide association (GWA) inferences on FE. A single-step genomic BLUP specification was adapted within both models to incorporate phenotypes on both genotyped and non-genotyped cows. To infer upon associations involving non-overlapping 1-Mb windows, we adapted a computationally efficient joint fixed effects test of all SNP markers in each window. Based on the use of a Bonferroni adjustment (Type I error rate = 5%) for the total number of such windows, significant associations with FE were detected within the same two windows using both the RFI and MT models. We also determined the number of genomic windows having significant associations with DMI, MBW, and MILKE to be 5, 5, and 3, respectively. None of these trait-specific windows overlapped with either of the two windows associated with FE, implying that FE is genetically conditionally independent of its component traits.

Key words: dairy cow, genome-wide association, feed efficiency, multiple trait analysis.

219

Genomic regions involved in seed protein, oil, and carbohydrate concentrations in glycine max

Ms Samantha McConaughy[1], Dr. George Graef[1]
[1]*University of Nebraska-Lincoln*

Soybeans are processed for their high-quality vegetable oil and protein meal for feed, food, and industrial applications but, because of the high negative correlations between seed protein and oil concentration, it has been difficult to develop soybean lines with concomitant increases in both protein and oil. Two soybean populations were developed using PI 547827 with lower total sugars as a common parent, crossed to two different soybean lines with modified protein and oil concentrations in the seed. The PI 547827 had seed protein concentration of 351 g kg^{-1}, with 197 g kg^{-1} oil. For the other two parental lines, one line had increased total oil concentration of 214 g kg^{-1} with 326 g kg^{-1} protein, and the second line had increased total protein concentration of 343 g kg^{-1} with 209 g kg^{-1} oil. The objectives of this study were to identify quantitative trait loci (QTL) related to seed protein, oil, and carbohydrate concentration as well as for the individual sugars sucrose, raffinose, and stachyose. For each of the two crosses, F4-derived recombinant inbred lines (RIL) were developed through single seed descent resulting in 526 and 404 RILs. Genotypes were determined for F4 plants by GbS, resulting in 4,000 to 6,000 polymorpphic SNPs used for QTL analysis. Populations were grown in an augmented design in two Nebraska and one Puerto Rico environment to produce seeds for compositional analysis by NIR. With 73% of lines in the high protein cross, and 21.4% of lines in the high oil cross exceeding processor targets 350 g kg-1 protein and 190 g kg-1 oil, these populations were successful in producing a large percentage of lines with improved seed composition, including transgressive segregation for all traits. The QTL analyses identified multiple QTL for protein, oil, carbohydrates, and each of the sugars on 13 different linkage groups.

220

Construction of genetic maps in complex autopolyploids

Marcelo Mollinari[1], Guilherme P Silva[1], Mitchell Schumann[2], Craig Yencho[3], Zhao-Bang Zeng[1], A. Augusto F. Garcia[4]

[1]Bioinformatics Research Center, Department of Statistics, North Carolina State University, [2]Department Soil and Crop Sciences, College of Agriculture and Life Sciences, Texas A&M University, [3]Department of Horticultural Science, North Carolina State University, [4]Genetics Department, "Luiz de Queiroz" College of Agriculture, University of São Paulo

Autopolyploid species play a fundamental role in agriculture. They are formed by multiple sets of chromosomes derived from the same species. This special condition imposes challenges to the construction of genetic linkage maps. Despite all advances in genetic mapping in autotetraploids, there is a lack of methods for construct genetic maps in organisms with ploidy levels greater than four, such as sweet potato, sugarcane, strawberry, some ornamental flowers and forage crops. The construction of these maps is a key step for genetic studies, including QTL analysis, assembly of reference genomes, and study of evolutionary process. In this work, we present a method and software to construct genetic linkage maps using multidose markers autopolyploids. The method combines exact multipoint-based models and two-point heuristics. Currently, the method is implemented for any even ploidy level up to 12 and the software is under development. To evaluate the method we conducted a simulation study comprising 500 autohexaploid full sib populations containing 250 individuals with one linkage group with 15 markers positioned at a fixed distance of 3 cM between them. Also, we analyzed a full sib tetraploid potato population dataset comprising 156 individuals and 5732 SNP markers. The results showed that in both scenarios the method perform reliably. In the vast majority of the simulations both the order and the linkage phase configuration of the markers were estimated correctly, with exceptions of inversions in closely linked markers. Also, the recombination fraction matrices obtained with the analysis of the potato dataset was highly monotonic, indicating a good estimation of the genetic map. Moreover, the multipoint method showed to be very important in complex scenarios where the two-point approach could not estimate the correct order and linkage phases.

221

Accuracy of genomewide prediction with and without heterozygote effects in apple

Elizabeth Blissett[1]
[1]*University of Minnesota*

The Rosaceae family includes such crops as apple, apricot, peach, and several other economically significant horticultural species. Within the Rosaceae family there is a growing demand to develop cultivars with superior horticultural quality and disease resistance. Horticultural quality not only includes traits like appearance, but also traits that contribute to consumer appeal such as crispness, juiciness, and flavor. The time and resources required to grow an apple seedling to maturity necessitates the use of genomic tools in predicting the performance of elite progeny. Genomewide prediction in self-pollinated species or in hybrid species such as maize (for which testcrosses with a common inbred are evaluated) does not require the modeling of heterozygote effects. Because apple is highly heterozygous, modeling heterozygote effects can potentially increase the prediction accuracy. We are assessing the prediction accuracy for apple crispness, firmness, and juiciness with clones from the University of Minnesota breeding program. Our preliminary results showed that including heterozygote effects neither increased nor decreased the prediction accuracy. We will be expanding our study to include additional traits and populations.

222

Investigating the genetic architecture of complex traits in a barley NAM population

Alex Ollhoff[1], Dr. Kevin Smith[1]
[1]*University Of Minnesota*

New genetic diversity must be incorporated into barley breeding programs if we are to efficiently meet market demands and mitigate the environmental impacts of agriculture. Germplasm collections are the best source of novel genetic diversity for plant breeding, but their accessibility is limited by our ability to phenotype many accessions of unadapted germplasm. To meet this challenge we, under the auspices of the Triticeae Coordinated Agriculture Project, created a nested association mapping (NAM) population bridging the genetic diversity in the National Small Grains Collection (NSGC) to barley breeding programs. The population was created by crossing 91 diverse parents from the NSGC Core Collection to the cultivar Rasmusson and creating ~80 recombinant inbred lines per cross. The population and founder parents have been characterized for plant height, bacterial leaf streak disease severity, and heading date. This population is also being characterized for SNPs using genotyping by sequencing. I will present an analysis of phenotypic variation for a number of quantitative traits in the barley NAM population. Improved plant genetic resources and analysis strategies will allow plant breeders to make more efficient progress toward a wider range of breeding objectives.

223

Genotype calling and linkage mapping from GBS data in Ipomoea trifida, a highly heterozygous species

Guilherme Da Silva Pereira[1], Marcelo Mollinari[1], Maria David[2], Federico Díaz[2], Wolfgang Grüneberg[2], Gabriel Margarido[3], Augusto Garcia[3], Bode Olukolu[1], Craig Yencho[1], Awais Khan[2], Zhao-Bang Zeng[1]
[1]North Carolina State University, [2]International Potato Center, [3]University of São Paulo

Genotyping-by-sequencing (GBS) has been applied in several species and broadly used in many genetic studies. However, there are still issues regarding genotype calling in highly heterozygous, outcrossing species. Here, GBS from 182 full-sibs of Ipomoea trifida (2n=2x=30) and their parents, M9 and M19, produced ~1.7 million tags from 516 million good, barcoded reads out of 811 million reads in total. Tags alignment rate was 74.69% when using an I. trifida draft genome as reference. A total of 38,959 polymorphisms (SNPs and Indels) were discovered and actual depths were stored in a variant call format (VCF) file using TASSEL-GBS pipeline. For genotype calling, we used a subjacent F1 segregation model implemented in SuperMASSA software to analyze 18,004 loci (>5 read depth). LOD score > 15 and maximum recombination fraction < 0.2 were used as grouping criteria in OneMap R package. From a total of 1,918 non-redundant, high confident markers (<25% missing data, >0.9 posterior probability), 1,692 markers were mapped in 15 linkage groups (LGs). LGs ranged from 65 to 200 markers and from 130.88 to 269.62 cM in length, spanning 2,595.94 cM in total (average marker density = 1.53). The LGs assembled 147 (0.48% of 30,377) scaffolds from the current draft genome, representing 276 Mb (59.66% of 462 Mb); 137 scaffolds were assigned to one LG each, nine scaffolds to two LGs and one scaffold to three LGs. Although big scaffolds could be uniquely assigned to LGs, assembling may need review due to these multiple assignments. This is an interesting example on how linkage map might assist genome assembling. Still, since I. trifida is a diploid wild sweet potato relative, its map will provide an important framework in furthering evolutionary studies between these species.

224

Performance of fixed vs. variable length haplotypes for genomic prediction

Melanie Hayr[1,2], Dr Tom Druet[3], Dr Andrew Hess[1], Dr Dorian Garrick[1,4]
[1]Iowa State University, [2]LIC, [3]University of Liege, [4]Massey University

Genomic prediction has traditionally been performed using models that fit covariates for SNPs; however, fitting covariates for haplotypes may improve prediction accuracy, or bias. Approximately 58,000 Holstein Friesian (HF), Jersey (Jer) and KiwiCross (KX) dairy cattle from New Zealand were genotyped on Illumina BovineSNP50 or HD panels. Genotypes at 37,740 SNPs were phased using LinkPHASE and DAGPHASE. Haplotypes were assigned using three methods: 1) Fixed length of 250 kb (Fixed); 2) Pairwise Linkage Disequilibrium (LD), where all SNP pairs in a haplotype must have D' ≥ 0.45; and 3) Recombination, where haplotypes ended when two neighboring SNPs had at least 20 recombinations. Genotyped females with milk fat records were separated into training (n = 23,907) and validation (n = 14,478) based on birthdates before or after 1 June 2008. Model performance was evaluated on the three breeds separately with accuracy and bias defined as the correlation or regression coefficient, respectively, of yield deviation on genomic breeding value. BayesA was run in GenSel fitting either SNP or haplotype allele dosage; haplotype alleles with frequency below 1% in the training dataset were filtered out of the analysis. Fixed and LD models resulted in higher accuracy than the SNP model for HF and KX (Fixed: 2.1% (HF) and 2.3% (KX) increase; LD: 1.3% (HF and KX) increase; $P<0.04$). No model was significantly different from any other for Jer ($P>0.20$). The Recombination model performed similarly to the SNP model for all breeds and was less accurate than the Fixed model for HF ($P=0.02$). Bias was not significantly different from one in any model. Fitting covariates for haplotypes can improve accuracy over fitting covariates for SNPs, however the method used to assign haplotype blocks is important. Haplotypes based on fixed length and pairwise LD performed similarly, but fixed length haplotypes take less time to generate.

225

Patterns of differentiation between house mouse subspecies at loci implicated in hybrid sterility: Do loci contributing to reproductive isolation show distinctive signatures in the genomic landscape?

Dr. Leslie M. Turner[1], Dr. Bettina Harr[1]
[1]*Max Planck Inst. For Evolutionary Biology*

A central goal in evolutionary biology is to understand the genetic basis of speciation. One strategy to identify loci involved in reproductive isolation is to compare the genomes of closely related species and identify highly differentiated candidate regions. Recently, this population-genomics approach has become widely adopted in studies of taxonomically diverse species pairs. An alternative approach is genetic mapping of reproductive isolation traits. Because mapping studies in natural populations are rare, it is not yet clear whether these two approaches will converge on the same candidate regions. We previously identified 53 loci associated with male sterility in a genome-wide association study of mice from a house mouse hybrid zone. Here, we compared genetic differentiation between subspecies in sterility loci to genome-wide averages to determine if they are outliers or show distinctive signatures revealing common features of their evolutionary history. We estimated sequence diversity and divergence (e.g., Pi, Fst, Dxy, Tajima's D) using genomes of eight individuals each from three Mus musculus domesticus populations (France, Germany, Iran) and two M. m. musculus populations (Kazakhstan, Czech Republic). We find that differentiation is higher, on average, in sterility loci, but these regions would not stand out in outlier scans. Hence, we cannot propose signatures which might be used to identify new candidate regions for reproductive isolation. However, measures of sequence differentiation will be useful for prioritizing candidate genes within sterility loci.

226

Alternative models for genetic analysis of tick and worm counts in Nellore cattle

Mr Tiago Luciano Passafaro[1], Mr Tom Wayne Murphy[1], Dr Fernando Brito Lopes[1], Dr Bruno Valente Dourado[1], Dr Fabio Luiz Buranelo Toral[2], Dr Guilherme Jordão de Magalhães Rosa[1]
[1]University Of Wisconsin, [2]Federal University of Minas Gerais

Ticks and worms are responsible for decreased productivity and subsequent economic losses in cattle populations worldwide. Traditionally, chemicals are the most common method to control parasites, but they can lead to increased parasite resistance and production costs. Consequently, alternative methods have been proposed to reduce the use of chemicals, including the selection of resistant animals. Statistical models for the analysis of tick and worm counts generally assume normality, which quite often is not a good approximation. Therefore, the objective of this study was to compare different generalized linear mixed models in the analysis of parasite resistance using data from a Nellore beef cattle herd. The data contained tick and worm counts collected at 550 days of age on 2,037 animals. Genetic parameters were estimated using Bayesian techniques. Gaussian and threshold models were used to analyze presence/absence of parasites, while Gaussian on original scale, Gaussian on logarithmic scale, and Poisson were used for the counting data. All models included contemporary group and age at recording as systematic effects in addition to the random additive genetic and residual effects. Spearman's correlation of predicted breeding values was used to compare model results. The traits displayed low to moderate heritability estimates ranging from 0.12 to 0.28 and from 0.06 to 0.23 for ticks and worms, respectively. Results indicate that variances in these traits are mainly accounted by environmental factors, but slow to moderate genetic improvement is certainly possible. Spearman's correlations between predicted breeding values obtained from the different models ranged from 0.38 to 0.99 and from 0.56 to 0.99, for ticks and worms, respectively. Results indicate a possible re-ranking of individuals depending upon the model used, especially between models considering the phenotype as binary on count variables. Therefore, model choice is an important component in the genetic improvement of resistance to ticks and worms.

227

Combining marker and pedigree information may enhance multiple-trait genome-enabled prediction

Dr. Mehdi Momen[2], Dr. Ahmmad Ayatollahi Mehrgardi[2], Dr. Fernando B. Lopes[1], Dr. Bruno D. Valente[1], Dr. Daniel Gianola[1], Dr. Guilherme J. M. Rosa[1]
[1]*University of Wisconsin-Madison,* [2]*Shahid Bahonar University of Kerman*

Multiple-trait prediction using molecular data exploits genomic correlations, and is perceived as potentially better than a classical pedigree-based multivariate analysis. Whether or not this holds, has not been explored extensively. Three tri-variate BLUP models were fitted using pedigree only (P), pedigree and markers (PM) or markers (M); PM treated the infinitesimal and genomic additive effects as independent. Traits were body weight (BW), breast meat (BM) and hen-house production (HHP) Aviagen, Inc., provided data on 1,351 chickens genotyped with a 600K Affymetrix chip. After quality control, 354,364 SNPs were retained. Phenotypes were corrected for several nuisance covariates and residuals were used as pseudo phenotypes. (Co)variance components were estimated via REML. Five-fold cross-validations were used to assess predictive ability with Pearson's correlation (r) and mean squared error (MSE) as metrics. Heritability estimates were 0.18±0.04 (BW), 0.24±0.05 (BM) and 0.31±0.07 (HHP) for P; 0.15±0.03, 0.24±0.04 and 0.17±0.04, respectively, for M, and 0.31±0.06, 0.41±0.05 and 0.36±0.07, respectively for PM. Clearly, PM captured a larger portion of the variation that using P or M alone. Genetic correlation between BW-BM, BW-HHP and BM-HHP were 0.48±0.13, -0.19±0.18 and -0.20±0.17 for P; 0.52±0.11, 0.03±0.2 and -0.20±0.17 for M, and 0.49±0.11, -0.03±0.19 and-0.17±0.17 for PM, respectively. Phenotypic correlations between traits were similar for all models; on average these correlations were 0.48±0.02, -0.06±0.03, -0.06±0.03 for BW-BM, BW-HHP and BM-HHP, respectively. For BW, the best predictive model was PM which had the largest r (0.22±0.04) and lowest MSE (46.36 ± 3.02). For BM, the M and PM models had a similar predictive ability (r: 1.96±0.04 vs. 1.95±0.05; MSE: 0.28±0.04 vs. 0.28±0.05 and). However, P-BLUP (r: 0.125±0.02) was worse than M-BLUP (0.24±0.02) and PM-BLUP (0.23±0.02) for HHP. Although there was some improvement in predictive ability using markers and pedigree together, the impact was trait-dependent. Perhaps further enhancement in predictive ability can be attained if pedigree and marker information were combined in a different manner.

Keywords: genetic correlation, genomic prediction, predict ability, heritability

228

The effect of PRRS virus disease outbreak on genetic parameters of reproductive performance in pigs

Mr. Austin M. Putz[1], Mr. Clint R. Schwab[2], Mrs. Alysta D. Sewell[2], Mr. Jack C. M. Dekkers[1]
[1]*Iowa State University,* [2]*The Maschhoffs*

Porcine reproductive and respiratory syndrome (PRRS) is the most costly infectious disease in the global swine industry. The objective of this study was to evaluate the effect of a natural PRRS outbreak in a reproductive farm on genetic parameters for four traits: number born alive, stillborn, mummified, and born dead. Data were collected on purebred Yorkshire and Landrace sows at three nucleus farms before (1319 Yorkshire and 1180 Landrace litters), during (635 and 607 litters), and after (928 and 921 litters) an outbreak of the 1-7-4 PRRS strain. Thirty-day rolling averages showed increases in stillborns and mummified piglets for approximately five months after the initial outbreak. Data were split into prior, during, and post phases using a semi-quantitative method based on a mixed model with a random farm-year-week effect. Genetic parameters for each trait across the three phases were estimated using bivariate animal models that included farm, year-month of farrowing, parity, and the 30-day rolling average of the trait as fixed effects and pedigree-based additive genetics of the sow as the random effect, along with a random permanent environment effect of the sow as needed. Heritability estimates ranged from 0.01 to 0.22 (SE 0.01 to 0.07) for both breeds and tended to be higher during the PRRS phase for numbers stillborn, dead, and mummified. Genetic correlations for the four traits between prior and during PRRS phases ranged from 0.48 to 0.63 (SE 0.36 to 0.82) in Landrace and from -0.79 to 0.85 (SE 0.12 to 0.87) in Yorkshire. Genetic correlations between prior and post phases ranged from 0.45 to 0.93 (SE 0.06 to 0.58) for both breeds. These results show an interaction between genetics and disease status for reproductive performance in pigs, suggesting that reproductive performance during a PRRS outbreak has a different genetic basis from reproductive performance without PRRS.

230

GxE models for performing predictions in complex scenarios: i) Tested genotypes in untested environments, and ii) Untested genotypes in untested environments

Diego Jarquin[1], Dr. Cinta Romay[2], Dr. Jode W. Edwards[3], Dr. Aaron J. Lorenz[4], Initiative The Genomes To Fields (G2F)[5]

[1]University Of Nebraska-Lincoln, [2]Institute for Genomic Diversity, Cornell University, [3]USDA-ARS Corn Insects and Crop Genetics Research Unit (CICGRU), [4]CFANS Agronomy/Plant Genetics UMN Twin Cities, [5]Consortium

Several genomic selection studies have been conducted in the context of multiple environments. However, only few of the prediction models deployed are flexible enough to include environmental information and a genotype-by-environment (GxE) component. Moreover, studies that have exploited multi-environment predictions, have mostly focused on only two prediction scenarios (CV2: predict tested genotypes in tested environments, and CV1: untested genotypes in tested environments) out of the four (additional: CV0: tested genotypes in untested environments, and CV00: untested genotypes in untested environments) that breeders might be interested from a practical point of view.
A first attempt to confront CV0 and CV00 schemes might consider the use of those models that work well for CV1 and CV2 and just predict genotypes for unobserved locations. Preliminary results using maize hybrid performance data from the multi-institution, multi-year Genomes to Fields (G2F) Initiative have shown average (0.53 and 0.33) predictive abilities across 24 locations for CV0 and CV00 schemes, respectively.
Since results were similar to those obtained with less elaborate models, the genomic, environmental and interaction components lose importance performing predictions in these scenarios.
For CV0 it was sufficient just train models using phenotypes (collected in other environments) of those genotypes being predicted while CV00 needed only SNPs.
As the predictions were null affected by the environmental conditions of target environments these values could be valid not only for the current environment but for other locations as well, something hard to believe.

We proposed a model that includes the interaction of genotype and environmental components, environment and environmental components, genotypes and environmental components and the three-way interaction of genotype, environment and environmental components. This model allows to perform oriented predictions for target unobserved environments such that these predictions become unique for a particular set of environmental conditions from a given site. The average correlation in 17 out of 24 locations was 0.44 for CV0, while for CV00 was 0.30 using 12 locations. Improvements might be expected with a better selection of the environmental components to be included. Here we have considered only means, minimum and maximum temperature, solar radiation, rain, and latitude and longitude location variables.

231

Modeling temporal estimates of genotype x environment interaction for Nile tilapia in three production systems

Arthur FA Fernandes[1], Bruno D Valente[1], Erika R Alvarenga[2], Eduardo M Turra[2], Guilherme J M Rosa[1]
[1]*Uinversity of Wisconsin Madison,* [2]*Universidade Federal de Minas Gerais*

Nile tilapia is the second most widely farmed fish worldwide, with a production of 3.67 million tons in 2014. It is produced in more than 100 countries under many different environmental conditions and production systems. Because fish are poikilothermic and sensitive to many environmental changes, the identification of genotype x environment interaction (GEI) that affects traits of interest is very important when devising breeding strategies to improve tilapia production. Therefore, the objective of the present work was to study GEI for body weight (BW) of tilapia produced under three alternative systems: cage, recirculating aquaculture (RAS), and biofloc technology (BFT). We performed the experiment in the aquaculture lab and the experimental farm of the Federal University of Minas Gerais, Brazil, with a total of 43 males and 86 females mated to produce 86 families. BW from 2,977 fish were measured at three times from 56 to 350 days of age approximately. Measures taken from fish in different production systems were treated as different traits and analyzed with a multi-trait random regression model (MTRRM) accounting for fixed effects of sex and group and random additive genetic, permanent and common family environment effects. One advantage of MTRRM is that (co)variance components between traits (production systems) can be assessed for any (combination of) time points. For example, the genetic correlation between BW at the 3 systems at the same age were initially 0.92 (BFTxRAS), 0.46 (BFTxcage) and 0.26 (RASxcage), and maximum values were achieved around 0.99 (BFTxRAS) at 60d; 0.90 (BFTxcage) and 0.74 (RASxcage), both around 200d. The results indicate that GEI effects are stronger between the cage and the other two systems, and that they increase with the fish age. This makes sense since the cage is the most distinct of the three systems, and also the most susceptible to weather conditions.

232

Construction of single-sample gene networks using transcriptomic and functional information for network QTL mapping

Ms. Qian Dong[1], Dr. Jack C M Dekkers[1]
[1]*Iowa State University*

To better understand regulatory genetic networks, methods have been developed to integrate genome-wide transcriptome data from a population sample of individuals with functional genetic information available on the species or across species. For example, Glass et al. (2013) developed PANDA (Passing Attributes between Networks for Data Assimilation) to integrate interactions between cooperating transcription factors and co-regulated genes by combining gene expression data, motif data and protein protein interaction (PPI) data. These methods build aggregate and consensus networks based on the gene expression data available across all evaluated individuals but do not show the potential heterogeneity in regulatory processes between individuals in a population. Recently, Kuijjer et al. (2015) developed LIONESS (Linear Interpolation to Obtain Network Estimates for Single Sample) as a method to derive single-sample networks by subtracting leave-one-out aggregate networks derived using PANDA from the whole sample aggregate network.
The objective of this study was to apply PANDA and LIONESS to repeated blood RNA-Seq data on pigs from an experimental challenge with the Porcine Reproductive and Respiratory Syndrome virus (PRRSv), combined with human motif and PPI data, to derive single-sample gene regulation networks in order to study the evolution of network topology over time and to characterize changes in gene regulation across and between individuals. The RNA-seq data used was obtained from globin-depleted whole blood samples collected 0, 4, 7, 10 and 14 days post infection with the PRRSv on eight littermate pairs of weaner pigs with genotype AA or AB at SNP marker WUR10000125, which has been shown to be strongly associated with weight gain and viral load. These data were analyzed separately for each time point.
Future work will use the single-sample networks and available 80k SNP genotypes to identify gene expression network quantitative trait loci by associating SNP genotypes with variations in edge weights for specific networks that summarize statistically relevant biological information. The hypothesis is that this will identify genetic variants that affect network regulation and capture essential biological processes that are relevant to a quantitative trait.

233

A shifting genetic architecture underlies selection response for flowering time adaptation in maize

Randall J. Wisser[1], Zhou Fang[2], Juliana E. C. Teixeira[3], John Dougherty[1], Teclemariam Weldekidan[1], Natalia de Leon[4], Sherry Flint-Garcia[5], Nick Lauter[6], Seth Murray[7], Wenwei Xu[7], Arnel Hallauer[8], Jim Holland[9]
[1]University of Delaware, [2]Bayer, [3]Empbrapa, [4]University of Wisconsin, [5]USDA-ARS, [6]USDA-ARS, [7]Texas A&M University, [8]Iowa State University, [9]USDA-ARS

The basis of response to artificial selection on a quantitative trait is well described in terms of allele frequency change, and methods to test for deviations from genetic drift are established. A weakness of inferring signatures of selection from allele frequency data is that these data cannot be directly related to a specific trait (consider multiple traits under selection). Therefore, we developed a study design to test for both marker-trait associations and deviations from drift in the same population at the same genomic variants. This method was applied to examine the genetic basis of tropical-to-temperate flowering time adaptation in maize achieved by a decade of phenotypic mass selection. Reference individuals from generations 0 to 10 (totaling 297) were sampled at random and self-pollinated to produce families that were evaluated in replicated trials at nine locations spanning nearly the full latitudinal range of the U.S. (Wisconsin to Puerto Rico). The reference individuals were genotyped at 55k SNPs, and marker-trait association were performed using family means as a proxy for the phenotype of the corresponding reference individual tests. Given the generationally stratified sample structure, the same genotype data was used to test for signatures of selection across generations. In addition, the study design enabled quantitative genetic inference on changes in genotypic means and genetic variances and analysis of genetic diversity, population structure and linkage disequilibrium. The results of this study imply a polygenic architecture underlying tropical-to-temperate adaptation in which the loci significantly associated with selection response change across generations. We also find intriguing evidence for the maintenance or increase in genetic variance in the presence of sustained mean response (phenotype) while also observing genetic hallmarks of selection (e.g. increases in local linkage disequilibrium).

234

Characterizing the allele-frequency spectrum of causal variants in yeast

Dr. Joshua S. Bloom[1,2], Laura Day[1,2], Holly Oates-Barker[1,2], Dr. Sebastian Treusch[3], Dr. Leonid Kruglyak[1,2]
[1]UCLA, [2]Howard Hughes Medical Institute, [3]Twist Bioscience

For many heritable traits, including common diseases in humans, the relative contribution of variants in different allele frequency classes to trait variation remains a subject of debate. We examined the genetic architectures of 41 quantitative traits in a whole-genome-sequenced panel of 13,500 yeast strains. The panel comprises 16 large biparental crosses among 16 diverse yeast strains,employing a round-robin design. We detected between 43 and 155 QTL per trait and explained most of the additive trait variance with detected QTL. We observe a large range of architectures across the different traits, with the rarest variants explaining an average of 60% of additive variance per trait (but ranging from 10% to 97% per trait). For many traits, a large fraction of trait variation is explained by many low-frequency variants with small effects. In addition, we observed widespread allelic heterogeneity within the mapped loci for some traits. Our highly powered design provides the clearest picture yet of the architecture of complex traits and the allele-frequency spectrum of causal variants, and informs the design and interpretation of future mapping studies.

235

Multiple-trait genomic selection methods to increase prediction accuracy in wheat quantity

Ms. Bettina Lado[1], Dr. Daniel Vazquez[2], Ms. Paula Silva[2], Dr. Martín Quincke[2], Dr. Lucía Gutiérrez[1,3]
[1]*Universidad De La República,* [2]*National Institute of Agriculture Research,* [3]*University of Wisconsin*

Genomic selection (GS) is a promising tool to increase the efficiency of selection in early stages of breeding programs. To apply GS at the beginning of breeding programs is necessary making predictions for multiple complex traits. The objective of this work was evaluate the accuracy of predictions using single and multiple traits prediction models. The analyses were performed using 412 wheat genotypes and four bread quality traits: dough strength (W), dough toughness (P), max height (MAXH) and SDS sedimentation volume. Approximately 100 lines were evaluated by years (2010 to 2013) with 17 checks in common between years. Genotyping by sequencing was performed and SNPs were discovered using TASSEL-GBS pipeline modified. Single traits prediction (STP) were conducted using GBLUP model in rrBLUP package in R software. Multiple traits predictions (MTP) were realized using kronecker product of correlation-traits-matrix and genotypic-relationship-matrix as variance-covariance matrix with GBLUP model too. Three cross-validation approaches were used to predict traits in randomly selected genotypes (40%). At first approach, model was training using the remaining genotypes (60%) and the prediction were performed by single trait (STP). In second approach, the remaining genotypes (60%) for four traits were used to train the MTP model. In third approach, the remaining genotypes (60%) for the predicted trait and all the genotypes for the other three traits were used to train the MTP. One hundred times cross validation was conducted. The correlations between adjusted phenotypes and genomic prediction using STP were high (0.383, 0.438, 0.453 and 0.458 for MAXH, SDS, W and P, respectively). With second approach, lower correlations were obtained in particular for MAXH and SDS (0.222, 0.209, 0.434 and 0.432, idem). In the last approach we obtained the highest correlations (0.577, 0.575, 0.479 and 0.595, idem). The third approach is useful to make predictions, in traits hard to measure, using correlated traits that were evaluated in testing genotypes too. However, to predict new genotypes without any phenotypic information is better to use a STP models.

236

Alternative approaches to modeling breeding value in tree breeding programs

Adam Festa[1,2], Dr. Lilian Matallana[1], Dr. Ross Whetten[1,2]
[1]*Department of Foresty & Environmental Resources North Carolina State University,* [2]*Functional Genomics Program North Carolina State University*

The goal of this study is to improve selection methods in tree breeding programs, by using patterns of gene expression and genetic variation in coding sequences to model parental breeding values. We hypothesize that early selection can be based on data from RNA-seq experiments on two conditions – first, that there are genetic differences among parent trees in gene regulatory networks, and second, that those differences are correlated with family mean performance in progeny field tests. Testing these conditions requires (1) obtaining reproducible estimates of gene expression from replicated samples of seedlings from OP, PMX, or CP families; (2) combining those family-mean estimates of gene expression levels into covariance estimates for parents that show predictive power in cross-validation studies for modeling phenotypic variation, and (3) combining covariance matrices based on coding sequence SNP variation, gene expression level variation, and pedigree-based estimates of allele sharing to optimize predictive modeling of phenotypic variation. A preliminary study is underway with a total of 62 different parent trees of Pinus taeda (loblolly pine), from a wide geographic distribution, with existing progeny field test data available from multiple sites. Seedlots (OP/PMX in 54 cases, CP in 8 cases) from these parents were grown in two different batches in a greenhouse, and pooled seedlings were harvested at age 3 months for RNA extraction, cDNA library preparation, and high-throughput sequencing. The reproducibility of family-mean gene expression patterns and the extent of differential gene expression have been assessed, and covariance matrices of gene expression and SNP variation are being used to model phenotypic variation in progeny field test data.

239

Identification of QTLs involved in quinolizidine and indole alkaloids content in Lupinus luteus L.

Dr. Claudia E. Osorio[1], Gustavo Del Canto[1], Anally Rupayan[1], Nicole Lichtin[1], Dr. Ivan Maureira-Butler[1]
[1]*Genomics and Bioinformatics Unit, Agriaquaculture Nutritional Genomic Center (CGNA).*

The generation of low alkaloid lupin varieties has facilitated their inclusion in human and animal nutrition and satisfied food industry standards. However, it has also increased susceptibility to herbivores and transmission of aphid-borne viruses and bacteria. Few studies about quantitative trait loci for alkaloids in yellow lupin have been conducted; thus, generating breeding tools to aid the efficient manipulation of theses secondary metabolites could significantly increase the generation of better nutritional quality lupin varieties.
To dissect the genetic architecture of quantitative alkaloid accumulation, we explored a sample of 164 diverse L. luteus accessions from several origins. The collection was assessed for alkaloid content for two consecutive seasons. Lupinine, sparteine and gramine were evaluated in two types of plant material, leaves and seeds. For association analysis, 315 co-dominant molecular markers (genomic SSRs, EST-SSRs, INDELs, and SNPs) were used to genotype the L. luteus accessions.
Using a mixed model (principal component analysis (PCA)+kinship matrix (K)), significant marker- trait associations were identified. Not surprisingly, associations between leaf tissues at two developmental stages showed variability between seasons, indicating a strong environmental influence on the production and storage of these secondary metabolites. Accumulation of alkaloids in seeds, on the other hand, was environmental stable for all three evaluated metabolites. Nine and eight markers were associated with lupinine and sparteine, respectively. Three markers showed associations with both lupinine and sparteine, suggesting one or more common QTLs for quinolizidine seed alkaloids. Gramine was associated with 11 markers. Three of them have been previously linked to a main QTL in a bi-parental population. To facilitate gene discovery, all significant markers are being fine mapped using backcross populations.

The Authors Acknowledge Financial Support From The National Commission For Scientific & Technological Research (FONDECYT Project 3140064), And The CONICYT Regional/GORE Araucanía/CGNA/R10C100.

240

Accuracy of estimated breeding values for hoof lesions in Canadian Holsteins

Francesca Malchiodi[1], Flavio S Schenkel[1], Anne-Marie Christen[2], David F Kelton[3], Filippo Miglior[1]
[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, [2]Valacta, [3]Department of Population Medicine, Ontario Veterinary College, University of Guelph

A key goal of dairy herds is to reduce the incidence of hoof lesions, which can be achieved both by improving management practices and through genetic selection. This study aimed to estimate genetic parameters for hoof lesions and to evaluate different approaches to increase the accuracy of the estimated breeding values (EBV). A total of 107,933 hoof lesions records from 53,654 animals were recorded by 23 hoof trimmers during routine trimming activities in 365 Canadian herds from 2009 to 2012. The lesions included in the analysis were: digital dermatitis, interdigital dermatitis, interdigital hyperplasia, sole hemorrhage, sole ulcer, toe ulcer, and white line lesion. Hoof lesions were coded as binary variables (0; 1), where 1 was assigned to the presence of a lesion. Univariate analyses were carried out considering either all of the observations available for a cow or including only the first hoof-trimming session per lactation. Subsequently, multi-trait linear analyses were performed. In addition to the presence of the lesion, locomotion (LOC) and the overall score for feet and legs (FL) were considered. Hoof lesions showed low heritabilities (from 0.01 to 0.07), but an exploitable genetic variation among sires was found. Using all of the observations available during the cow's lactation increased the accuracy of EBVs. The reliability of EBV for hygiene-related lesions also increased when LOC and FL were included in the multi-trait evaluation. Therefore, the use of all hoof lesion records available for a cow and indicator traits can increase the reliability of EBVs for hoof lesions, but the improvement was moderate. A further improvement in the breeding value accuracy could be achieved using genomic data, which will be investigated next.

241

The genetic basis of the coordination between resource allocation and resource acquisition in Drosophila melanogaster

Dr. Elizabeth King[1], Enoch Ngoma, Anna Perinchery, Patrick Stanley
[1]*University of Missouri*

Given a finite amount of resources, a central requirement of all organisms is optimizing the allocation of those resources to competing anatomical structures and physiological functions, which necessitates the coordination of multiple organ systems within the nutritional environment. Despite the central importance of resource allocation, we know very little how variation at the genome level translates to variation in resource allocation and ultimately variation in high-level phenotypes such as aging, obesity, and reproduction. We utilize a large multiparental mapping population, the Drosophila Synthetic Population Resource to genetically characterize a suite of interrelated allocation phenotypes across multiple levels of the genotype to phenotype map in multiple nutritional environments. We identify the loci that influence how resources are partitioned differently when resource availability varies and characterize the patterns of co-localization between traits and between diets.

242

Studying the genetics of microRNA expression for alcohol related traits [in a panel of recombinant inbred mouse strains]

Dr Pratyaydipta Rudra[1], Dr Wen J. Shi[1], Brian Vesta[1], Pamela Russell[1], Dr Boris Tabakoff[1], Dr Paula Hoffman[1], Dr Laura Saba[1], Dr Katerina Kechris[1]
[1]*University of Colorado, Anschutz Medical Campus*

MicroRNAs (miRNAs) are small non-coding RNAs that bind messenger RNAs and promote transcript degradation or repress translation. We conducted a high-throughput investigation to discover brain miRNA that influence predisposition to alcohol-related phenotypes in a panel of recombinant inbred (RI) mouse strains. The LXS RI panel is derived from two inbred progenitor strains (Inbred Long Sleep and Inbred Short Sleep), which were selectively bred for differences in sensitivity to the hypnotic effects of an acute dose of alcohol. Using existing genotype data on the panel and newly generated miRNA sequencing data, we explored the genetics of miRNA expression. To investigate what proportion of miRNA expression is attributed to genetic variation, we developed a statistical framework to measure and test heritability of the miRNA expression. We also developed a novel statistical method for efficiently finding miRNA expression quantitative trait loci (eQTL) affecting their expression. The eQTL were compared with known alcohol related QTL found by prior studies. Several miRNA eQTL and high heritability miRNA were identified and some of them were from the same region of the genome associated with alcohol phenotypes of interest, including drinking patterns and low dose ethanol activation. We found that brain miRNA expression can be a highly heritable molecular trait in mouse. Furthermore, we identified a number of loci responsible for miRNA expression that are also associated with alcohol phenotypes. These loci and miRNA are important candidates for understanding the genetic predisposition of the phenotypes modulated through miRNA mechanisms.

Funded by NIH/R01 AA021131

243

Comparing genotyping methods for development of genomic selection prediction models in Eucalyptus

Rodrigo F. Santos[1], Braulio F. X. Moraes[2], Alexandre Missiaggia[3], Aurélio M. Aguiar[3], Bruno M. Lima[3], Donizete C. Dias[3], Gabriel D. P. S. Rezende[3], Flávia M. A. Gonçalves[2], Márcio Resende[4], Patricio R. Munoz[1], Matias Kirst[1]

[1]University of Florida, [2]Federal University of Lavras, [3]Fibria Celulose, [4]RAPiD Genomics

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. Advances in next-generation sequencing and methods of genome complexity reduction, such as sequence-capture, allow the development of flexible genotyping platforms, 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 evaluated the impact of sequence-capture and array-based genotyping methodologies on the development of GS prediction models for a Eucalyptus breeding population composed of 739 trees, phenotyped for 13 wood quality and growth traits. The descriptive analysis of the datasets detected differences in linkage disequilibrium patterns, minor allele frequency, missing data and marker distribution between both genotyping methods. RR-Blup and BayesB prediction models developed for both genotypic datasets resulted in similar predictive abilities, estimated using cross validation, to evaluate the performance of GS models. This demonstrates that both genotyping methodologies are equivalents for development of GS models, on the traits and population evaluated.

244

Use of environmental convariates for genomic selection modeling genotype by environment interaction in a wheat bredding program

Mr. Pablo Martín González Barrios[1], Mrs. Bettina Lado[1], Dr. Martin Quincke[2], Mrs. Paula Silva[2], Dr Marcos Malosetti[3], Dr Lucia Gutiérrez[4]

[1]Facultad De Agromomia -Universidad De La República , [2]Instituto Nacional de Investigación Agropecuaria, [3]Wageningen University, [4]University of Wisconsin

The study of genotype by environment interaction (GEI) is a key factor when evaluating phenotypic performance of genotypes in multi-environment trials. Actually, it is necessary to capture the GEI information within genomic prediction models for the aim of improving the prediction accuracy and selection efficiency of superior cultivars adapted to specific environmental conditions. The use of environmental covariates has shown positive results in the characterization of environments and the study of adaptive responses to environmental stress variables. However, it is unclear in the context of genomic selection the best strategy to select the most relevant environmental covariates and how include this information in prediction models. The objective of this study was to evaluate strategies incorporating selected environmental covariates to modeling GEI in the context of genomic selection in a wheat breeding program. Information from the Uruguayan National Wheat Breeding Program of five years (2010-2014) was used. Yield evaluation of 74 elite wheat lines in the 19 year-sowing date combinations were used and genotyped using genotyping by sequencing. Twenty four environmental covariates using the most relevant meteorological variables for vegetative, flowering and grain filling periods were calculated for each genotype, sowing date and year. By using factorial regression the covariates with greater phenotypic effect on performance across different environments were selected. Mixed models were used to predict by-environment (GBLUPgxe), using an equal environmental correlation with environmental covariates (Cov_gxe) and using and environmental correlation matrix with environmental covariates as fixed effect (Cov) for different sets of environments. Predictions were made for each year using information from the remaining years of evaluation. No differences in prediction accuracy among models were found in 2010. However, in the remaining years the model Cov showed significant higher predictions. The inclusion of environmental covariates improves the predictive ability of genomic prediction models in predictions by years.

245

Post-hoc blocking and genotype by environment interaction in Zossiagrass

Lin Xing[1], Kevin Kenworthy[1], Salvador Gezan[2], Bryan Unruh[3], Patricio Munoz[1]
[1]*Agronomy Department, University of Florida,* [2]*School of Forestry Resources and Conservation, University of Florida,* [3]*Environmental Horticulture, University of Florida*

The Randomized complete block design (RCBD) is the most widely used experimental design in biology, including turfgrass breeding programs, due to its simplicity and decent isolation of environmental variation. In early stages of breeding programs, plant breeders screen many genotypes to identify the best performing to be used a parents. However, in RCBD as the number of breeding lines increases, the block size increases, which can cause the blocks to loss their original designated function, controlling the environmental variation. Thus, the estimation of parameters and prediction of breeding values is negatively impacted from the uncontrolled environmental variation.

To improve the analysis accuracy, our study incorporated spatial factors in the analysis by post-hoc blocking. Post-hoc blocking involves superimposing a different blocking structure onto the original experimental design. Post-hoc blocking was tested using a series of five experiments with 80 genotypes of zoysiagrass (Zoysia spp.) established across Florida. Two post-hoc blocking designs were evaluated (incomplete block design and row-column design), and their model fitting and residual variances were compared with the original RCBD.

Results indicated that the row-column design was consistently superior to the original RCBD design in terms of residual variances and data fitting. The incomplete block design performed either slightly better or similar to the original RCBD design. Most importantly, the increase in prediction accuracy was accompanied by a change in the rank of breeding values of evaluated genotypes. After the determination on the model selection for the data analysis, we evaluated different genetic parameters, such as h2 and the partition and understanding of the sources of genotype by environment (GXE) interaction for zoysiagrass. Together, these results can have a significant impact in the selection of experimental design and provided important parameter information for zoysiagrass cultivar development.

246

Super panel development and genomic analyzes using two high-density SNP platforms in Nellore beef cattle

Dr. Ricardo Ventura[2], Dr. Flavio S. Schenkel[1], Dr. Stephen P. Miller[3], Dr. Miguel H. Santana[6], Mr. Gerson A. Oliveira Junior[6], Dr. Minos E. Carvalho[6], Ms. Elisangela C. Mattos[6], Dr. Gordon Vander Voort[2], Dr. Fabyano F. Silva[4], Mr. Pablo A.S. Fonseca[5], Dr. Jose B. Ferraz[6]
[1]University of Guelph, [2]BIO, [3]AgResearch, [4]Federal University of Vicosa, [5]Federal University of Minas Gerais, [6]University of Sao Paulo

The use of genomic information to assess genetic variation has been a revolution to animal breeding in the last few years. Several panels containing different number of SNPs from two main companies have been released to date. This research investigated two high-density panels (The Affymetrix Axiom Genome-Wide BOS1 Array (648,874 SNP) and the Illumina High-Density BovineHD Bead Chip Array (777,962 SNP)), its combination to create a super panel (SP) containing 1,261,128 SNPs, and five low and medium commercial panels defined as subsets of the BovineHD (6K, 20K, 30K, 50K, 74K). The marker distribution according to the chromosome location was first studied to identify gaps across different panels. Runs of Homozygosity (ROH), imputation and linkage disequilibrium (LD) pruning analyzes were performed after genotyping a Nellore beef cattle population (N=279) with both HD platforms (BOS1 and BovineHD). After applying a k-fold validation process, the overall average imputation accuracy (concordance rate) per animal from BovineHD and BOS1 to SP was 97.75% and 98.48%, respectively. For the same scenarios, imputation accuracy per SNP, measured as a squared Pearson correlation, was also performed. BOS1 slightly outperformed BovineHD, influenced mainly by regions with reduced minor allele frequency. The SP strategy reduced the gap size in some chromosomes (i.e.: Chrs 6, 14 and 23) and did not improve uncovered regions such as in Chrs 7 and 27. LD pruning, applied to remove markers in high LD (r2 > 0.9), kept 25% more markers in the pruned BOS1 data set, compared with the pruned BovineHD set. ROH results before pruning showed an average number of segments in homozygosity per animal slightly higher (4%) for BovineHD and an average segment size (KB) slightly larger for BOS1 (2.6%). The super panel combining SNPs from both platforms allowed coverage improvement for several chromosome regions. Further investigation including new analyzes, such as Copy Number Variation (CNV), is required before making decision on the HD panel genotyping strategy for Nellore cattle.

247

Applying quantitative genetics advances towards molecular breeding in rice.

Dr. Michael J. Thomson[1]
[1]*Texas A&M University*

Tremendous progress has been made in the molecular dissection of quantitative traits in crops, yet few breeding programs have fully integrated these advances. QTL mapping and cloning provide in-depth knowledge on key genes controlling complex traits, while genome-wide association studies highlight beneficial alleles across diverse sets of germplasm. At the same time, next-generation sequencing and high-throughput SNP genotyping have accelerated the rate of discovery and enabled more efficient marker-assisted selection. As one of the world's major food crops, rice has benefited from a wealth of QTL and association studies—but more work is needed to fully realize the potential of molecular breeding tools. During my time at the International Rice Research Institute, we set up the Genotyping Services Lab to assist with diversity analysis, fingerprinting, and QTL mapping using a 6K SNP chip developed at Cornell University, while trait-specific SNP markers were validated and deployed for major genes and large-effect QTLs. At the same time, the 3,000 Rice Genomes Project provided an excellent opportunity to identify SNP haplotypes at genes needed for rice improvement. After joining Texas A&M University last year, my program is now focused on implementing large-scale allele mining, SNP validation, and marker deployment in rice. One initiative is to compile and share trait-predictive SNP haplotype and marker validation information across different germplasm groups, which will enable breeders to access the most relevant SNP markers for their programs. A community effort and online resource called RiceSNPvisor is proposed to provide rice breeders with information and advice on breeding-relevant SNP haplotypes, through compiling published data on SNP associations and guiding future efforts at validating SNP haplotypes. The ultimate objective is to enable rice breeders to more readily employ high-throughput SNP genotype data and imputation to track targeted SNP haplotypes at key genes across their mainstream breeding pipelines.

248

Hybrid prediction effectiveness using genome-wide information in tobacco

Dr. Bruna Carvalho[1], Dr. Carlos E Pulcinelli[1], Dr. Ramsey S Lewis[2], Dr. Adriano T Bruzi[3], Dr. Magno A P Ramalho[3]

[1]British American Tobacco, [2]North Carolina State University, [3]Universidade Federal de Lavras

Identifying high performing hybrids is one of the most essential and expensive part of breeding programs, especially for self-pollinated species. Because of that, one of the benefits vaunted in the use of genomic selection (GS) is the ability to predict hybrid performance from parental line genetic information without the need to evaluate these individuals in the field. Moreover, the number of possible hybrid combinations is already large with a relatively small number of lines. GS would allow identifying promising hybrids, when themselves and other hybrids produced from their parents were not tested in field trials. In tobacco, because of complexity in evaluating genotypes phenotypically, it is not possible to test a large number of hybrids in the field. Thus, the possibility of using GS for this crop is an enormous advantage. We aimed to verify if it is possible to predict hybrids of Flue-Cured Virginia tobacco using GS. For that, 13 high performing lines were sequenced by GBS and 68 hybrids from their crossings where evaluated in two environments using triple lattice design. Plots consisted of 10 plants, with 1.15m between-row and 0.45m between-plants spacing. Traits assessed were: yield (YLD), general quality index (GQI), steam percentage (STM), total sugar content (SUG), total alkaloid content (ALK) and an additive linear index considering all these traits (INDEX). For hybrid prediction, the SNP value of each hybrid was determined by combining SNP values from their respective parental. Marker effects were estimated using G-BLUP method and only parental information were used for training the model. Accuracies in predicting hybrids were determined by the correlation between estimated genomic value and observed phenotypic information. Correlations between genetic divergence of lines (based on markers) and hybrid performance, and heterosis were also checked. The predictive accuracy of hybrids ranged from 0.47 (SUG) to 0.69 (ALK). The correlation between genetic divergence among lines and the hybrid performance, and heterosis was null. These results have shown that using GS cannot be decisive for selection of hybrids, instead it can be very useful to reduce the number of hybrids assessed in the field, eliminating those combinations with less probability of success.

249

Performance of genomic prediction for a sugarcane commercial breeding program

Itaraju Brum[1], Mark E Sorrells[1]
[1]*Cornell University*

Sugarcane is a crop of economic importance in tropical areas and mostly used for production of sugar, ethanol, energy and animal feed. Modern Sugarcane varieties are hybrids between two autopolyploid species, the domesticated “noble cane” Saccharum officinarum (2n=80) and the wild Saccharum spontaneum (2n=40-128). Sugarcane breeding comprises a series of steps for effective evaluation of clonally multiplied plants, a process that can take more than a decade. Most relevant traits, e.g. biomass and sucrose content, are quantitatively inherited. In this context Genomic Selection is being studied as a tool to increase efficiency in the breeding program.

A population consisting of 1876 clones from two sugarcane breeding cycles at CTC (Sugarcane Technology Center, Brazil) was genotyped by sequencing probes captured as randomly generated DNA fragments and analyzed as Single Nucleotide Polymorphisms (SNP). An initial set of 245k SNPs was filtered for adequate read depth, heterozygosity and missing data resulting in a final set of 30k SNPs. Alignments of markers to the genome of sorghum, sugarcane's closest relative with a sequenced genome, shows extensive genome coverage. Analysis of association between molecular markers also shows linkage disequilibrium extending long distances within chromosomes and among chromosomes.

This population was phenotyped for plot weight, Brix, fiber and Pol (sucrose content), with replicated measurements taken on two harvests (plant crop and ratoon crop). Broad-sense heritabilities for those traits ranged from 0.37 to 0.90. Genomic selection was then evaluated through the accuracy of cross validation within a breeding cycle and between cycles, resulting in accuracies ranging from 0.25 to 0.62 in the former case, and 0.0 to 0.30 in the latter. Here we observed a strong genotype by year interaction effect leading to reduced accuracies when predicting across breeding cycles. This effect, coupled with moderate accuracies in cross validation, need to be taken into account in the deployment of Genomic Selection for a sugarcane breeding program.

250

Trans regulatory architecture of genetic transcriptome variation from 1,000 yeast individuals

Dr. Frank W. Albert[1,2], Dr. Joshua S. Bloom[2], Dr. Jake Siegel[2], Laura Day[2], Dr. Leonid Kruglyak[2,3]
[1]*University of Minnesota,* [2]*University of California,* [3]*Howard Hughes Medical Institute*

* FWA and JSB contributed equally

"Expression quantitative trait loci" (eQTL) are an important source of genetic variation for many traits. To date, all eQTL studies have been hampered by low statistical power due to limited sample sizes. We have addressed this limitation in the yeast Saccharomyces cerevisiae by sequencing transcriptomes in more than 1,000 recombinant individuals generated from a cross between two yeast isolates. The statistical power of this dataset permits mapping of thousands of previously "missing" eQTL that together account for ~80% of the heritability of gene expression.

A typical transcript is influenced by a median of 6 and up to 20 eQTL. While 43% of all genes had a local eQTL close to the gene, most eQTL are located elsewhere in the genome and influence gene expression in trans. Rather than appearing randomly across the genome, the vast majority of the newly discovered trans eQTL fell into one of 111 hotspot regions that each affect the expression of up to thousands of genes. By combining information from all genes that map to a given hotspot, we can fine-map the causal hotspot location with high precision, in 26 cases to single-gene resolution of less than 1.5 kb. This permits a first systematic analysis of the types of genes that act as trans eQTL. Finally, despite our high statistical power, many local eQTL did not act as trans eQTL for other genes. This might indicate that the expression changes these eQTL cause at their local genes do not further affect cellular physiology, at least not in ways that are reflected in the transcriptome. This raises important questions about the characteristics of genes where local regulatory variation does have cellular and phenotypic consequences.

251

Functional SNP in a polygenic disease induced by high-altitude in fattening Angus steers by combining multi-OMICS data into Systems Biology

Dr. Angela Canovas[1], Rebecca R. Cockrum[2], Dale Brown[3], Suzette Riddle[3], Joseph Neary[2], Tim Holt[2], Greta M. Krafsur[2], Juan F. Medrano[4], Alma Islas-Trejo[4], Mark Enns[2], Scott Speidel[2], Kristi Cammack[2], Kurt R. Stenmark[3], Milton G. Thomas[2]
[1]University Of Guelph, [2]Colorado State University, [3]University of Colorado-Denver, [4]University of California-Davis

High-altitude (>1800m) disease is a challenging problem in beef and dairy cattle. The disease is consequential of hypoxia-induced right ventricular (RV) heart failure as per vascular inflammation of the pulmonary artery (PA) and hypertension. The disease has moderate to high heritability ranging from 0.2 to 0.4; however, minimal information exists of the genes involved.The transcriptomes of six tissues (i.e. left and right ventricle, pulmonary artery, aorta, muscle, and lung) were examined in samples harvested from fattening-yearling Angus steers phenotyped to be of low or high pulmonary arterial pressures (LPAP and HPAP; n = 10/group). Tissue specific splice variants were identified in RV (n=555), aorta (n=547) and PA (n=152; $p<0.01$ FC>2) between LPAP and HPAP animals. Most of the splice variants are located in key regulators genes with roles in angiogenesis and cardiomyopathy (NFATC1), movement of leukocytes and neutrophils (OLR1, PLAUR), failure of heart (CTGF), hypertrophy of heart ventricle (TREM1, GATA2) and vascularization (SYVN1). Since high-altitude disease is a polygenic disease and single-SNP within a positional-candidate gene explains limited variation, there is great interest in discovery functional SNP located, for example, within nodes and hubs of gene networks. Various “omics” tools such as transcriptomics combined with gene networks and systems biology assist with discovery of coding and tag SNP. Therefore, several SNP variants segregated specifically in either the LPAP or HPAP animals have been identified. Among them, 139 SNP were located in key regulator genes involved in the adaptation of the disease.These approaches helped identify splice variants corresponding to key regulator genes in a polygenic disease induced by high-altitude in Angus cattle. The discover of functional SNP associated with high-altitude disease by combining structural and functional genomic data into a systems biology approach will help to develop more robust approaches for genetic selection in beef cattle providing the opportunity to enhance EPD accuracy.

252

Impact of relatedness on genomic prediction and GWAS detection in two elite eucalyptus breeding populations

Ms. Bárbara Müller[1], Leandro Neves[2], Janeo de Almeida Filho[3], Marcio Resende Jr[2], Annette Fahrenkrog[3], Patricio Munoz[4], Matias Kirst[3], Dario Grattapaglia[5]

[1] University of Brasília, [2]RAPiD Genomics LLC, [3]University of Florida, School of Forest Resources and Conservation, Forest Genomics Laboratory, [4]University of Florida, Agronomy Department, [5]EMBRAPA Genetic Resources and Biotechnology

Genomic Selection (GS) and Genome-Wide Association Studies (GWAS) are promising approaches to accelerate breeding cycles and increase selection intensity in forest trees. We report results of GS and GWAS for growth traits in two real-life, closed breeding populations of Eucalyptus benthamii (BEN, n=505) and Eucalyptus pellita (PEL, n=732), using a 60K SNP chip (EUChip60K). Predictive abilities (rgy) were 0.15 and 0.42 for BEN and PEL respectively, corroborating the limited genetic diversity available in BEN due to a severe bottleneck during its introduction in Brazil. Genomic prediction models were fitted using SNPs on each chromosome separately to evaluate their individual contribution to rgy and genomic heritability. The rgy using individual chromosomes were close to the genome-wide estimate, suggesting that prediction is driven largely by relatedness between training and validation. Conversely, genome-wide heritability estimates were higher than single chromosomes ones, indicating that genome-wide coverage maximizes capturing additive genetic variance. To assess the relative impact of relatedness versus LD (Linkage Disequilibrium) on rgy, models were fitted minimizing relatedness between training and validation, based on Principal Component Analysis (PCA) results. As a control, cross-validation was carried out using random allocation of individuals to training and validation. As hypothesized, rgy obtained by minimizing relatedness were zeroed for BEN and approximately half of the rgy achieved in the control for PEL. This result suggests that for PEL some LD is captured into predictions although relatedness is likely the main driver. GWAS identified 120 SNPs associated with volume for PEL before correcting for relatedness. When the genomic relationship matrix was included in the model, only one significant association was detected for a SNP located in an exon of a gene involved in cellulose biosynthesis. This study corroborates the key role of relatedness in the practice of GS and the limited value of GWAS in operational breeding populations.

253

Modeling genotype by environment interaction over multiple covariates for feed efficiency in dairy cattle

Yongfang Lu[1], Dr Michael J. Vandehaar[1], Dr. Diane M. Spurlock[2], Dr. Kent A. Weigel[3], Dr. Louis E. Armentano[3], Dr. Charles R. Staples[4], Dr. Erin E. Connor[5], Dr. Mark D. Hanigan[6], Dr. Zhiquan Wang[7], Dr. Robert J. Tempelman[1]

[1]Michigan State University, [2]Iowa State University, [3]University of Wisconsin-Madison, [4]University of Florida, [5]USDA Animal Genomics and Improvement Laboratory, [6]Virgina Tech University, [7]University of Alberta

Genome wide association (GWA) studies based on high-density genomic markers are important to filter genomic regions, and eventually identify candidate genes, associated with traits of interest such as feed efficiency (FE). However, most GWA studies infer an overall marginal effect for each marker across environments, without allowing for potential interactions between markers with environmental covariates. Marker by environment interaction (MxE), as a proxy for genotype by environment (GxE) interaction, is plausible for a complex trait like FE, particularly in our dairy consortium of seven research stations managed over a wide range of climatic and production systems. Additionally, it seems that multiple environmental covariates could determine GxE within a reaction norm (RN) modeling framework; however, as the number of such covariates increases, such inferences can become intractable. Instead of directly modeling MxE effects separately for each covariate and marker, we propose a RN extension to a genomic animal model (GBLUP) whereby collective animal by environment (AxE) effects are fitted such that the dimension of the mixed model equations does not exceed that of GBLUP. We also present a strategy to backsolve for marker- and covariate-specific MxE effects from AxE effects. Based on simulation embedded on the genotypes and design structure of our consortium data, we determined that our model had superior GWA properties to GBLUP. We also applied both models to the dairy consortium data, specifying linear, quadratic, cubic, and quartic terms on three environmental covariates: average contemporary group production, relative humidity, and maximum temperature. Production was the most important environmental covariate contributing to GxE, being 6% of the magnitude of the genetic variance for FE. In a cross-validation study, our proposed model yielded 3% greater prediction accuracy compared to a regular GBLUP analysis.

254

Working on Evolvix, the first general-purpose programming language designed by biologist for biologists.

Dr. Laurence Loewe[1]
[1]*University of Wisconsin-Madison*

To be efficient, programmers like to use the “right language for the job”: C for speed, R for statistics, etc. But which language is right for biology?

Answers either target a special approach (recommending specialized languages like SBML) or focus on the execution of algorithms (recommending general-purpose languages with appropriate features). Specialized languages are very efficient at solving *their* special problems, but can be prohibitively difficult to use, even for “closely related” problems, as they lack the general-purpose programming features that would enable more efficiency by breaking restrictive assumptions made by these special languages.

General languages make less assumptions and thus allow code to be more flexible when the specific assumptions of useful specialized languages cannot be shared. Such expressivity can come at the price of allowing for dangerous code that greatly complicates finding errors if left unchecked by compilers.

Biology’s extraordinary diversity regularly breaks the assumptions of existing general purpose programming languages. For example, biologists regularly use diverse synonyms, ontologies, alignments, tables, missing data, uncertainty, probabilities and other concepts not easily implemented from scratch.

Evolvix began as a special purpose model description language aiming to make it easy to accurately model biological systems (starting with pure mass action systems, current prototype). Using Evolvix for modeling showed the need for general-purpose programming features. Without these features it will not be possible to achieve the ultimate goal of Evolvix, that is to support evolutionary systems biology by simulating reliable, testable, interactive overviews of nestable, dynamic fitness landscapes that mechanistically predict how genotypes and environments affect fitness and how populations evolve. Simplifying the removal of the many numerical, statistical, logical, computational, data-related and other errors that invariably plague complex code in biology, requires the rigorous reduction of inessential programming complexity and is at the heart of the development strategy for Evolvix.

255

Investigating kinship and heterotic relationships of expired maize plant variety protection lines

Mr. Mike White[1], Dr. Mona Mazaheri[1], Mr. Joseph Gage[1], Dr. Mark Mikel[2], Dr. Natalia de Leon[1], Dr. Shawn Kaeppler[1]
[1]University of Wisconsin, [2]University of Illinois

The Plant Variety Protection (PVP) Act of 1970 gives breeders exclusive right to distinct sexually or tuber propagated plant varieties for a period of 20 years. Upon expiration of the certificate, PVP maize inbred lines become publicly available and are released through the North Central Regional Plant Introduction Station. Expired-PVP maize inbred lines can be an excellent source of elite germplasm to enhance or initiate a breeding program. The use of expired-PVP maize inbred lines in a breeding program is challenging because pedigrees are often ambiguous, making it difficult to utilize lines in appropriate heterotic pools. Empirical observations in the literature and from our group indicate that the concept of heterotic groups is much more complex than the canonical Stiff Stalk vs non Stiff Stalk grouping. We evaluated 240 expired-PVP and public reference inbred lines at 430K SNPs derived from RNA-sequencing. Relationships of lines were determined using neighbor joining and visualized in a circular tree dendrogram. To further assess relationships of expired-PVP lines, an admixture for K groups was performed and identity by state matrices were generated to visualize diversity within this set. Using cross validation scores, K= 8 yielded the best fit for the expired-PVP panel. The admixture results placed a traditional founder in each of the eight groups. These include; B14, B37, B73, Oh43, Mo17, PH207, PHWG5, and PHGG7. Finally, a chromosome identity analysis was preformed to finely visualize chromosomal parental contributions to progeny. Together, these analyses provide conclusions on the similarities and exclusivities among all heterotic pools and between private sector breeding programs. Utilization of relatedness information of expired-PVP lines can help to better categorize lines with previously unknown pedigree information to expedite the utilization of expired-PVP material in a breeding program.

256

Joint local testing of variant-sets improves power and interpretation of gene-by-context interactions

Mr. Francesco Paolo Casale[1], Dr. Oliver Stegle[1]
[1]*European Bioinformatics Institute*

Understanding the genetic architecture of complex traits remains a central question in quantitative genetics. One of the major avenues is to understand how contextual variables, such as environmental factors, impact the genetic basis of complex traits. To address this, we here present a new statistical approach based on linear mixed models (iSet) that i) jointly models the effect of multiple variant from a genetic region and ii) accounts for categorical context variables. iSet generalizes a wide class of previous mixed models and scales to large cohorts.

We show that iSet offers substantial power advantages compared to previous interaction tests. Additionally, the model differentiates between classical interactions with context-specific effect sizes and complex cases where the configuration of causal variants changes across the contexts. Finally, we show how iSet leads itself to a generalization of narrow sense heritability estimates using mixed models, allowing to partition phenotypic variance into persistent genetic effects, and different classes of gene-by-context interactions.

When applying iSet to a monocyte stimulus eQTL study, which includes naive and stimulated cell states, iSet yielded a 54% power increase to detect interactions compared to state-of-the-art methods and additionally revealed more than one thousand cases where gene-by-context interactions are due to heterogeneity in causal variants. We characterized these eQTLs, revealing surprisingly rich genetic architectures, frequently involving multiple independent effects. Additionally, we report new insights into the mechanisms that underlie eQTLs with directional changes across contexts.

iSet can also be used to perform interaction analysis in stratified sample, which we explored by performing a gene-by-sex interaction analysis of lipid traits in individuals from the Northern Finland Birth Cohort. Despite the moderate sample size (N=5,256), our analysis reveals genome-wide significant genotype-context interaction with sex. We validate these observations using large meta-analyses, where the reported interactions replicate at nominal significance levels.

259

Phenotypic variation and genetic dissection of silage yield and compositional traits in recombinant inbred testcrosses in Maize. (Zea mays L.)

Jonathan Renk[1], Calli Anibas[1], Joseph Gage[1], Shawn Kaeppler[1], Natalia de Leon[1]
[1]*Department of Agronomy, University of Wisconsin-Madison*

Maize silage provides many farmers producing or raising livestock with a high energy forage. This is particularly important in states like Wisconsin where close to 20% of the total maize acreage harvested is dedicated to silage production (USDA, 2015). The goal of this study was to conduct a genetic dissection of forage yield and key compositional traits for 516 maize recombinant inbred lines (RILs) derived from three different populations [149 intermated B73 x Mo17 (IBM), 183 Oh43 x W64a (OWRI), and 184 Ny281 x H99 (NyH)] evaluated as testcrosses using inbred PHG47 as the tester. The RIL testcrosses were evaluated in a replicated field trial across three environments in 2012 and 2013 in South Central Wisconsin. Forage dry matter yield ranged from 2.064 to 11.5 ton/acre for IBM, 3.173 to 10.53 ton/acre for OWRI, 3.078 to 10.85 ton/acre for NyH and percentage of dry matter biomass ranged from 25.82 to 61.35% for IBM, 31.73 to 56.77% for OWRI, and 32.92 to 59.32% for NyH. Among the most relevant compositional characteristics, percentage of crude protein ranged from 6.84 to 11.25% for IBM, 6.37 to 10.55% for OWRI, and 6.42 to 10.5% for NyH. Mean values of those traits were relatively similar between populations. A quantitative trait loci (QTL) for percentage of starch content was identified in chromosome 1 for IBM. QTL for cell wall digestibility and percent crude protein were identified in chromosome 1 and 6 respectively for OWRI. The identification of genomic regions associated with these traits could help in the development of superior maize silage varieties.

Source: USDA. Crop production 2014 summary. 2015.

Funding acknowledgement: United States Department of Agriculture (USDA) Hatch WIS01639

260

Breeding strategies and multivariate approaches to optimize drought resistance in plants

Dr. Eliane Bodah[1], Dr. Bruce Weir[1]
[1]*University of Washington*

Drought stress has become one of the major concerns worldwide due to climate changes. Breeding for adaptable crops, that can maintain satisfactory grain yield while withstanding stress, is a crucial area of research. Drought resistance has been reported as a quantitative trait influenced by the environment. In this sense, we present three approaches that can be used to elucidate how drought resistance and/or tolerance might be investigated in plants. First, we developed a rapid morphological protocol under greenhouse conditions to aid phenotyping. Second, we applied this protocol to a wide germplasm screening, and conducted crosses between susceptible and tolerant pea varieties. Populations from these crosses are being advanced for future mapping. In these first two approaches, plants were grown in randomized completed blocks with four replicates in a greenhouse located in Moscow, ID. Plants were subjected to optimal irrigation (1400 ml as a control) and two treatments that reduced optimal irrigation rates by 40% (840 ml, moderate drought) and 60% (560 ml, severe drought) to induce drought stress. The tests were repeated. In all cases, varieties significantly ($P<0.05$) differed in their response to water deficiency. Similarly, there were significant ($P<0.001$) differences among treatments on growth parameters, with highest values recorded at optimal irrigation. Third, with the aid of a QTL mapping dataset generated by Pauli et al. (2016) we proposed integration of multivariate approaches in marker-trait association studies, and tested their feasibility for drought resistance. In this third approach, we utilized partial least squares regression and Simes extended procedures to help incorporate multiple traits to marker analysis of the QTLs associated with stress-response cotton.

262

A hyperparameterized whole genome simulator for developing and evaluating genetic designs for population and quantitative genetic studies

John J Dougherty[1], Doctor Randall J Wisser[1]
[1]*University of Delaware*

In population and quantitative genetics simulators are useful for exploring hypotheses about the basis for evolutionary change, assessing the efficiency of experimental designs and determining the properties of test statistics. Using the core functionality of simuPOP (a forward-time simulation environment) we have developed a flexible, whole genome simulator that can be extensively parameterized based on prior information on trait genetic architectures and is capable of implementing a range of breeding schemes. Our aim is to create and evaluate new classes of genetic study designs and test statistics that provide optimal power and potential for inference. In our initial work, we are examining a study design to study the genetics of selection response that allows a genome-wide association study (GWAS) and a signature of selection study (SoSS) to be performed on the same population. In this design, families from multiple generations of selection are genotyped and also evaluated together in one or more environments. The resulting data are then used for GWAS-SoSS and associated loci can be examined in terms of trait-specific allelic effects and corresponding allele frequency changes. Starting with real genotype data on inbred lines of maize we simulated the construction of a multi-parental synthetic population. Under a genetic architecture defined from prior studies on flowering time in maize, this population was then intermated for multiple generations with and without selection. The simulation output were then piped into statistical testing frameworks for GWAS and SoSS. This presentation will provide an overview of the simulator with baseline performance evaluation and examination of the benefits and limitations of such a GWAS-SoSS study design.

263

Statistical methods for the functional characterization of selection signatures

Jaon Lapeyronnie[1], Claude Chevalet[1], Dr. Bertrand Servin[1]
[1]*GenPhySE, Université de Toulouse, INRA*

Understanding processes that shaped population diversity in the past can help in predicting evolution. For example, inferring the evolution of genetic polymorphism in response to past adaptive constraints can help anticipating genetic response to future changes in the environment. The detection of genomic regions that responded to selection in the past, sometimes called "selection signatures", can be based on genetic data alone, or exploit additional information on individual or populations. Here we present the extensions of two models based on genetic data alone (FLK [1] and hapFLK [2]) to include covariates and exploit phenotypic or environmental information on populations. Based on simulations, we show to what extent including covariates while exploiting linkage disequilibrium information can improve detection power compared to other approaches. Finally we show how our simulations can allow to predict the efficacy of different sampling strategies to study genome response to adaptation.

Keywords : (Population genetic), selection, covariates, statistics

[1] Bonhomme, Maxime et al. “Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended.” Genetics 186.1 (2010): 241–262. PMC. Web. 19 Feb. 2016.
[2] Fariello, María Inés et al. “Detecting Signatures of Selection Through Haplotype Differentiation Among Hierarchically Structured Populations.”Genetics 193.3 (2013): 929–941. PMC. Web. 19 Feb. 2016.

264

Genomics-aided development of Cucumis Hystrix introgression - Library for cucumber improvement

Paradee Thammapichai[1], Yonghua Han[2], Tao Zhang[1], Jiming Jiang[1], Yiqun Weng[1]
[1]*University of Wisconsin-Madison,* [2]*Jiangsu Normal University*

Cucumis hystrix (2n = 2x = 24, HH) is the only known species that is cross-compatible with cucumber and has a great potential for cucumber improvement. To facilitate introgression of C. hystrix chromatins into cucumber genetic background and to understand the mechanisms of genetic recombination between these two species, de novo draft genome assembly of C. hystrix and an oligonucleotide-based chromosome painting technique were developed. The genomes of two C. hystrix accessions, TH1 and CN1, were sequenced with Illumina and 454 technologies and a draft genome was assembled which contained 16,865 scaffolds with the largest scaffold being 467 kb and N50 scaffold size of 23.8 kb. The estimated genome size of C. hystrix was ~447 Mb. Draft genome assemblies of CN1 and TH1 were used for comparative analysis with the melon and cucumber genomes, and development of polymorphic markers for genetic mapping in C. hystrix as well as design of probes in oligo painting. Oligos specific to a single chromosome of cucumber were identified and massively synthesized de novo. The synthesized oligos were amplified and labeled with biotin or digoxigenin for use in FISH. These probes produced FISH signals on a single cucumber chromosome and were used to paint homoeologous chromosomes in other Cucumis species. We were able to precisely map the pairing between cucumber chromosome 7 and chromosome 1 of C. hystrix in an F1 hybrid. These two homoeologous chromosomes paired in 71% of prophase I cells but only 25% metaphase I cells, which may provide an explanation of the higher recombination rates compared to the chiasma frequencies between homoeologous chromosomes reported in plant hybrids. Using SSR markers, over 100 ILs carrying varying number and size of C. hystrix chromosomes or fragments have been molecularly characterized.

266

Fine-mapping a major maize domestication QTL for ear diameter

Mrs. Alessandra York[1], Dr. John F Doebley[1]
[1]*University Of Wisconsin-Madison*

Maize ears are much larger in diameter and have more rows of grain than maize's ancestor, teosinte. This significant difference in ear structure makes it an essential trait to study to better understand domestication. Previously, our lab mapped domestication quantitative trait loci (QTL) in a set of maize-teosinte hybrid recombinant inbred lines (RILs) and identified a large effect QTL on the short arm of chromosome 5 for ear diameter. This QTL co-localized with a large effect QTL for kernel row number, and represents a likely domestication target. Following this work, an unsuccessful effort was made to identify the gene underlying this QTL by fine-mapping with a BC_2S_3 heterogenous inbred family (HIF) derived from the RILs. Our lab also mapped this QTL using an independent set of a maize-teosinte BC_6S_6 lines segregating for this QTL region. With this set of lines, the QTL was confirmed and more narrowly mapped to a ~2.654 Mbp region. We are now attempting to fine-map the QTL using three of the BC_6S_6 lines to create two families that segregate for the QTL. In these two families, the QTL is located between two marker loci which includes 54 candidate genes. We used these markers to identify recombination events in the region and generated a set of 163 recombinant chromosome nearly isogenic lines (RCNILs). More markers within the 4.81 cM region are being identified to fine-map where the breakpoints are in each of the RCNILs. Once the phenotyping data has been collected in the upcoming summer, we'll be able to identify a causal region or gene that is responsible for part of the difference in ear diameter between maize and teosinte. When the gene is identified, we can identify the causative variant within the gene, and test whether selection during domestication has reduced diversity at this gene.

267

Characterization of the genetic architecture of maize ancestor, Teosinte

Chin Jian Yang[1], Alessandra M. York[1], Michael R. Tuholski[1], Wei Xue[1], Weidong Wang[1], Lora L. Daskalska[1], Michael A. Neumeyer[1], Bode A. Olukolu[2], Jinliang Yang[3], Vince S. Buffalo[3], Jeffrey C. Glaubitz[4], Peter J. Bradbury[5], Sharon E. Mitchell[4], Edward S. Buckler[5], Jeffrey Ross-Ibarra[3], James B. Holland[6], John F. Doebley[1]

[1]Laboratory of Genetics, University Of Wisconsin-Madison, [2]Department of Crop Science, North Carolina State University, [3]Department of Plant Sciences, University of California, [4]Institute for Genomic Diversity, Cornell University, [5]USDA-ARS, Cornell University, [6]USDA-ARS, Plant Science Research Unit, North Carolina State University

Maize was domesticated from teosinte in the Balsas river valley at about 9000 years ago. Previous studies have identified several key genes that distinguish maize from teosinte, including teosinte branched 1 (tb1), teosinte glume architecture 1 (tga1), and grassy tillers 1 (gt1). Interestingly, most of these major domestication genes were selected from standing variations that are still present in modern teosinte populations. This suggests that modern teosinte populations can be useful in understanding the selection process during domestication, as well as providing a resource for modern maize breeding. Here, we attempt to characterize the genetic architecture underlying domestication related traits in a population of teosinte from Palmar Chico, Mexico. We are undertaking a two-part approach here: (1) estimate variance parameters and genetic correlations as an understanding on how selection could have acted on a single trait and multiple traits (2) identify quantitative trait loci (QTL) from genome-wide association studies (GWAS). Phenotypic data collection is now complete for ~4800 plants from two seasons (winter 2013 and 2014) of field trials in Homestead, FL. Genotypic data is obtained using genotype-by-sequencing (GBS) method. Data analyses including genotype imputation and mixed linear models (MLM) fitting for variance parameters estimate and GWAS are currently ongoing.

268

Heritability and genetic gain of Bipolaris resistance in switchgrass

Mr. Kittikun Songsomboon[1], Mrs. Jamie Crawford[1], Mrs. Jaime Cummings[2], Dr. Gary Bergstrom[2], Dr. Donald Viands[1]

[1]Plant Breeding and Genetics Section, School of Integrative Plant Science, Cornell University, [2]Plant Pathology and Plant-Microbe Biology Section, School of Integrative Plant Science, Cornell University

Switchgrass (Panicum virgatum L.), a selected biomass crop, suffers from seed rot and leaf spot caused by Bipolaris oryzae (Breda de Haan) Shoemaker resulting in poor establishment and yield reduction. A sustainable approach is to improve disease resistance in switchgrass germplasm. Prior to breeding, narrow-sense heritability of leaf spot resistance based on replicated half-sib progenies in 'Kanlow' was estimated to determine genetic variability for resistance as well as an effective breeding method. However, the heritability was not significantly different from zero, indicating either that the genetic variability of disease resistance might be low in this population and/or that improvements need to be made in the screening technique. Recurrent phenotypic selection has been completed for the first cycle in 'Cave-in-Rock' and 'Shelter' for resistance to either seed rot and leaf spot. The gain of selection for leaf spot resistance in 'Cave-in-Rock' in the first cycle was low at only 2.5% with low realized heritability at 0.03. On the other hand, seed rot resistance was significantly improved both in 'Cave-in-Rock' and 'Shelter' from 26.86 ± 2.77% to -7.29 ± 4.20% and from 20.99 ± 4.81% to 1.62 ± 3.85% of germination reduction, respectively. More research needs to be done to determine how to make progress from selection for resistance to the leaf disease.

269

Genetic architecture of the maize inflorescence shattering trait

Michael Tuholski[1], Weidong Wang[2], Chin Jian Yang[1], Zhihong Lang[3], John Doebley[1]
[1]University of Wisconsin-Madison, [2]Purdue University, [3]Chinese Academy of Agricultural Sciences

The domestication of modern cereal crops from wild grasses provides a powerful system for studying rapid phenotypic evolution. Wild grasses shed their seeds at maturity to ensure propagation but this scattering or 'shattering' of seed would have been detrimental to cereal grain harvest by humans. The loss of natural seed dispersal would therefore have been key in domestication. In sorghum, rice, and foxtail millet, this shattering trait has been mapped to the Sh1 locus which encodes a YABBY transcription factor. Maize shattering QTL track to regions on chromosomes one and five containing Sh1 orthologs (ZmSh1-1.1 and ZmSh1-5.1+ZmSh1-5.2). Thus this project seeks to confirm the genetic architecture of the modern maize non-shattering trait on two levels. First, a QTL fine-mapping experiment will confirm that ZmSh1-5.1+ZmSh1-5.2 is the causative gene and will determine the nature of the change in the gene which led to non-shattering maize from its shattering ancestor teosinte. Preliminary results using developed recombinant inbred lines suggest the causative region lies in the 5' UTR of ZmSh1-5.1+ZmSh1-5.2. Second, this project will examine the interaction of the chromosomes one and five maize genes (ZmSh1-1 and ZmSh1-5.1+ZmSh1-5.2) to observe any additive effects between the two genes. Preliminary results suggest a less-than-additive relationship although further analysis is ongoing.

270

Enhancement of genomic prediction and QTL-mapping in Bovine Respiratory Disease using sequence imputation and feature selection

Jesse Hoff[1], Jared E. Decker[1,2], Robert D. Schnabel[1,2], Jeremy F. Taylor[1]
[1]*Division of Animal Sciences, University of Missouri,* [2]*Informatics Institute, University of Missouri*

Genomic Selection and QTL mapping have proven to be very successful approaches for research and breed improvement. Expanding the application of Genomic Selection to traits outside those routinely measured but with high economic importance such as feed efficiency and disease resistance is an area in which these technologies hold promise. There have been limits to the translation of results from experimental models to industry applications, due to issues that include genetic distance and statistical power. Utilization of imputed sequence level variation data can improve the yield of existing data sets by refining the genetic signal and better describing genomic relatedness accounting for rare variation. In this study of Bovine Respiratory Disease Complex, 2703 case-control Holstein animals sampled from two locations with Illumina BovineHD genotypes were imputed to the 1000 Bull Genome Project Run 5 reference set using FImpute after including variants for an additional 30 sequenced animals from the study population. Over 11 million variants were found with minor allele frequencies of at least 5%. These variants used to fit linear mixed models in GEMMA that increased chip heritability from 13.6% to 14.4% over the BovineHD data. There was also a substantial increase in the resolution of QTL regions. These imputed variants can be used in conjunction with functional annotations of BRDC networks based on RNA-seq challenge experiments as well as the functional annotation of genome itself. These approaches are being implemented to improve genomic predictions based on imputed genotypes from training populations. For traits like BRD with low heritability and that are also difficult to measure, maximizing the value of genomic discovery populations is critical for trait improvement. Moving from genomewide markers to features selected from accurate sequence level imputation can transition Genomic Selection from black box associations to a more precise biological understanding of complex diseases.

271

Biochemical analysis of epicuticular wax in onions: A Thrips resistance strategy

Eduardo D. Munaiz[1], Dr. Michael Havey
[1]*University of Wisconsin-Madison*

Thrips (Thrips tabacci) is the main insect pest of onion and can cause 50% of yield loss and its control is highly dependent on insecticides. Thrips is also the principal vector of Iris yellow spot virus which can cause over 40% yield loss. Lower amounts of epicuticular waxes on the leaves has been associated with fewer thrips, less feeding damage, and lower incidence of IYSV. The aim of this study is to study the chemical composition of epicuticular waxes on an array of 17 plant introductions (PI) in order to assess natural variation for the amounts and compositions of specific wax components that may show resistance to thrips. The PIs were grown in four environments and leaf samples evaluated using gas chromatography-mass spectrometry (GC-MS). Results showed that visual phenotypic scores correlated with the wax composition and amounts. Hentriocontanone-16 and octaconasol-1, the prevalent waxes in wild-type onions, were lower in lighter-green colored (semi-glossy) types. Low wax (glossy) phenotypes showed the lowest amount of hentriocontanone-16, but can possess larger amounts of other waxes. Glossy and semi-glossy phenotypes showed lower numbers of thrips relative to waxy phenotypes. This study supports the effectiveness of the biochemical analysis of epicuticular waxes in order to identify the unique wax profiles associated with thrips resistance.

272

GWAS for resistance to stem rot and aggregated sheath spot of rice

Juan Rosas[1], Victoria Bonnecarrere[1], Sebastian Martinez[1], Fernando Perez de Vida[1], Gaston Quero[2], Shubert Fernandez[1], Silvia Garaycochea[1], Jean-Luc Jannink[3], Lucia Gutierrez[4]
[1]INIA Uruguay, Rice Research Program, [2]College of Agriculture, University of the Republic, [3]Cornell University, [4]University of Wisconsin

Stem Rot and Aggregated Sheath Spot are among the major diseases affecting temperate rice worldwide, and are caused by the fungi Sclerotium oryzae (SCL) and Rhizoctonia oryzae-sativae (ROS), respectively. Resistance to these diseases is quantitatively inherited and has low heritabilities in field trials, making conventional breeding difficult. Furthermore, reports on QTL for resistance to both diseases are scarce. Thus, identification of QTL is needed for marker assisted breeding for these traits. We analyzed the association between resistance to both diseases in 643 Uruguayan rice advanced inbred lines (327 indica and 316 tropical japonica ssp.) and two sets of GBS SNPs (49.6K for indica and 28.9K for japonica). Resistance was measured in four years of field trials in Eastern Uruguay and in two (for SCL) and three (for ROS) greenhouse trials with a 0-9 scale. Phenotypic means were spatially and phenologically corrected, and weighted based on each trial heritability. Two mixed models, one with P (PCA scores) for japonica, and another with P+K (kinship) matrices for indica, were used for GWAS accounting for different levels of relatedness. Trial heritabilities for ROS were 0.06-0.43 in field and 0.67-0.89 in greenhouse, while for SCL ranged were 0.24-0.72 in field and 0.65-0.76 in greenhouse. Consistent QTL across environments and/or diseases were found in both subpopulations: 24 in indica and 8 in tropical japonica, with allelic effects from 0.4 to 2.0 points in the 0-9 disease scale. Although some favourable alleles of these QTL were also associated with plant height (up to +5cm) and flowering time (up to +4.5 days), 16 QTL in indica and 2 in tropical japonica had specific effects for resistance to the diseases. These findings provide useful tools for molecular breeding of disease resistance in advanced temperate rice germplasm.

273

Fine mapping coincident QTL for multiple traits on 6H of barley - The use of traditional method to resolve unfavorable associations of quantitative traits

Lu Yin[1], Ahmad Sallam[1], Ed Schiefelbein[1], Guillermo Velasquez[1], Karen Beaubien[1], Kevin Smith[1]
[1]*University of Minnesota*

Barley with bright kernels (low KD level), low deoxynivalenol (DON), resistance to Fusarium head blight (FHB), net blotch and bacterial leaf streak (BLS), relatively low grain protein (GP) content (115 - 135 g/kg), and good agronomic traits (high yield, low lodging, low stem breakage, desirable heading and senescence dates) are desirable for malting. Quantitative trait loci (QTL) studies and other genetic evidence have associated these traits with a region of chromosome 6H near bin 6 of barley. Yet it remains to be known the basis of associations of these trait at this locus. For example, it is unknown at this locus whether high GP somehow makes the plant more resistant to KD (pleiotropic effect), or GP and KD are conferred by independent, tightly linked loci. A population was developed, consisting of recombinant near isogenic lines (rNILs) in the background of elite cultivar Lacey and different introgression of Chevron spanning the 6H QTL region. There were 1968 F2 lines, and 269 F4 lines. Previous studies narrowed down the causal region for DON, KD and GP to a 1.1 cM interval, but have not been able to determine the association basis of these traits. In this study, a larger set of F4 rNIL lines are to be phenotyped and genotyped with genotype-by-sequencing which gives denser marker coverage. The above mentioned traits have been measured in field trials in RCBD at two locations, and NB at additional greenhouse trials. Phenotypic analysis shows significant line effects for senescence, FHB, heading date, yield, GP, KD, and a non-significant line effect for BLS. These results suggest that BLS is unlikely to be associated with the 6H QTL region while confirms the other traits are associated with the region. Some traits, however, show mixed results, requiring further investigation. Genotypic analysis is still in progress.

274

Shared selective sweeps across human populations

Kelsey E Johnson[1], Benjamin F Voight[2,3]
[1]*Genetics and Gene Regulation Graduate Program, Perelman School of Medicine, University of Pennsylvania,* [2]*Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania,* [3]*Department of Genetics, Perelman School of Medicine, University of Pennsylvania*

Positive selection has been a driving force in human evolutionary history, and genome sequencing data from diverse populations now allows the examination of this process at unprecedented detail. Signals of positive selection can be shared across populations, such as the lactase persistence haplotype across European populations; or unique to an individual population, like high altitude-adaptive mutations in Tibetans. The degree to which signals of positive selection are unique vs. shared across human populations remains unexplored, and thus our understanding of the relative importance of local vs. global adaptations in recent human evolution is incomplete. We scanned 20 populations from the 1000 Genomes Project for evidence of positive selection, specifically recent hard selective sweeps, using the iHS method. We identified putatively selected regions that overlap across two or more populations, and performed simulations to determine if the number of overlaps across populations and continental groups corresponds to their evolutionary relationships. We found significantly more sweep overlaps than expected across continental groups, and within all continental groups except Africa. We classified shared sweeps from overlaps by identifying those overlaps with the same sweeping haplotype and with their best sweep-tagging SNPs in strong linkage disequilibrium. We identified pervasive sweep sharing, with more closely related populations sharing a larger proportion of sweeps, but including shared sweeps across continental groups. To identify potential differences in the biological pathways targeted in shared sweeps, we are exploring enrichment of GWAS SNPs in these genomic intervals using the tool GREGOR. In conclusion, we found evidence for shared sweeps in human populations, including across continents, suggesting a common evolutionary origin for adaptive variants in diverse populations. Future work will focus on identifying the biological targets of shared sweeps and resolving the evolutionary scenarios leading to shared sweeps.

275

Dissecting the genetic architecture of local adaptation using environmental association and selection mapping in soybean

Mr. Nonoy Bandillo[1], Dr. Justin Anderson[2], Dr. Michael Kantar[3], Dr. Robert Stupar[2], Dr. James Specht[1], Dr. George Graef[1], Dr. Aaron Lorenz[2]
[1]University of Nebraska-Lincoln, [2]University of Minnesota, [3]University of Hawaii

Local adaptation is critical for crop species in the face of climate and environmental change but its underlying genetic causes is not clearly elucidated. This study aims to identify candidate genes associated for environmental and climatic adaptation as well as identify selection signals important for modern improvement globally and locally. In this study, we used the Glycine max accessions maintained in the U.S. Department of Agriculture Soybean Germplasm Collection and genotyped with the SoySNP50K genotyping platform. Using FST analysis between elite and landrace populations, we identified signatures of selection in America, China, Japan and Korea. Signatures of selection were observed on every chromosome, with notable regions identified for E genes known to control soybean flowering time and maturity. Overall, the pattern of selection signals showed that accessions collected in Japan are diverged from those collected in China and America. To identify loci underlying variation for local adaptation, we performed genome-wide association (GWA) mapping on 112 environmental variables that were collected based on latitude and longitude coordinates of 3,012 geographically diverse landrace accessions. Environmental variables included information on temperature, precipitation, and soil chemistry and texture, as well as altitude data. GWA mapping identified important candidate loci that putatively contribute to adaptation to environmental stresses. Interestingly, some of the associated loci identified from GWAS had been differentially selected within countries. Using these results, we will identify loci that contributed strongly to fitness variation within G. max and explore possible genetic mechanisms that may facilitate a flexible response to a changing environment.

276

A widely-applicable and powerful method for detecting balancing selection

Katherine M. Siewert[1], Dr. Benjamin F. Voight[2,3]
[1]*Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania ,* [2]*Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania,* [3]*Department of Genetics, Perelman School of Medicine, University of Pennsylvania*

Balancing selection preserves two or more alleles at a locus over long evolutionary time periods. This type of selection is known to play a role in human evolution, a classic example being the Hemoglobin Beta locus, which maintains alleles that protect against malaria, but cause sickle-cell anemia. Owing to large-scale genomic data sets, discovery and characterization of loci subject to balancing selection has become an active area of investigation. Our goal is to quantify the number of sites and determine which genes are subject to selection in a wide-range of species. To answer these questions, a test must be powerful, robust to assumptions about demographic history that are difficult to model with certainty, not rely on outgroup sequences and be computationally efficient. We are developing an approach that utilizes the signature of allelic class build-up, in which neutral variants accumulate at the same frequency as the balanced allele they are linked to. Preliminary results show that our method performs comparably to more computationally intensive methods, despite our statistic not requiring knowledge of complex demographic parameters or outgroup sequences. This indicates our method can be used with a wide range of species. We will next apply our method to human and drosophila population data to detect loci under selection.

277

Vernalization, dormancy, and the annualization of onion (Allium cepa) for breeding

Christopher J. D'Angelo[1], Dr. Irwin L. Goldman[1]
[1]*University of Wisconsin-Madison*

The biennial life cycle of onion (Allium cepa) provides a significant challenge for plant breeders. The lengthy vernalization period, coupled with bulb dormancy in some varieties, mandates a two year generation time from seed to seed. The goal of this research is two-fold: 1) to characterize the requirements for vernalization, dormancy, and floral induction in cultivated onion, and 2) to develop a reliable protocol for breeders that will allow for onion seed to be produced on an annual cycle in a winter greenhouse while still offering a bulb evaluation.

A time course experiment conducted over two years with F1 hybrid onion bulbs (Cortland & Sherman, Bejo Seeds) held at 10°C was used to determine the minimum vernalization time required for floral competency. Bulbs were removed from cold storage at two week intervals, planted in a greenhouse at 15°C, and monitored for sprouting, scape emergence, and flowering. The date of each developmental change was recorded for individual bulbs.

The optimum cold treatment to satisfy the vernalization requirement and overcome dormancy for Cortland and Sherman bulbs was found to be between 14 and 16 weeks at 10°C. The average times to scape emergence in the greenhouse for these vernalization lengths were 85 and 74 days post-chilling, respectively. In the 2013-2014 season, field grown onion bulbs flowered and produced viable seed in a greenhouse within approximately 13 months. This seed was planted for the 2014 field season and is the first instance of approximating an annual cycle for onions in our breeding program. This work has laid the foundation for future experiments with qPCR and RNA sequencing of the doubled haploid line CUDH 2107 to identify candidate genes responsible for vernalization, dormancy, and floral development.

278

Optimizing the training population composition to improve genomic prediction accuracy across selection cycles of a barley breeding population

Tyler Tiede[1], Ed Schiefelbein[1], Karen Beaubien[1], Guillermo Velasquez[1], Dr. Shiaoman Chao[2], Dr. Kevin P. Smith[1]
[1]*University of Minnesota,* [2]*USDA-ARS*

Genomic selection (GS) is becoming routinely used in plant breeding. To date, the literature and anecdotal experience of users indicates that, while it works, the accuracy of GS is greatly influenced by the composition and size of the training population (TP). The University of Minnesota barley breeding program has been using and evaluating GS since 2009. Now, a balanced experimental dataset including 50 progeny selected via GS and 50 progeny selected at random from each of the first three cycles of selection has been evaluated in multiple trials for grain yield and deoxynivalenol concentration, the toxin produced by the fungal pathogen Fusarium graminearium. Using this dataset to estimate the true breeding values of selection candidates in each breeding cycle, we retroactively assess the effect that three TP updating and optimization strategies have on prediction accuracy, which include i) amending the original TP to include breeding lines to more comprehensively represent the selection candidates, ii) updating the TP with data collected on lines from previous breeding cycles before predicting subsequent cycles, and iii) enacting algorithms proposed to identify optimal TP subsets.

279

Fine-mapping of QTL associated with genotype-dependent plant regeneration in maize

Mr. Frank McFarland[1]
[1]*University of Wisconsin-Madison*

The most widely used and efficient crop genetic engineering protocols currently require generation of embryogenic, regenerable callus cultures as part of the transformation process. The ability of crop plants to produce embryogenic, regenerable cultures and regenerate plants is, however, highly genotype dependent, significantly hindering progress in crop genetic research and enhancement. We are conducting research aimed at deciphering the genetic mechanisms controlling differential embryogenic, regenerable response in tissue culture, which would aid in the development of improved, genotype-independent plant tissue culture and transformation systems. An inbred backcross (IB) introgression mapping population was developed from the cross of inbred maize line A188 (high embryogenic, regenerable response) to maize inbred line B73 (low embryogenic, regenerable response). The recurrent parent in the backcrosses, B73, is widely utilized in maize genetics research and has undergone genome sequencing. Using this initial mapping population, several QTL with significant effects on embryogenic, regenerable response were identified, with the QTL having the largest effect located on the long arm of chromosome 3. Initial fine mapping of the QTL region on chromosome 3 has been conducted via molecular marker and phenotypic analysis of lines derived from further backcrosses to B73. Currently, a genomic region 2cM in length and another of 10cM in length, located within the QTL on chromosome 3, have been shown to be associated with high embryogenic, regenerable response. Further fine-mapping population development and screening is underway to identify recombinants with reduced genomic segments that still express high embryogenic, regenerable tissue culture response. Ultimately, the goal of this project is to identify and characterize the underlying genetic factor(s) controlling genotype-dependent embryogenic, regeneration response, and to potentially produce a novel germplasm that maintains the A188-derived response in a B73 genetic background.

282

The construction and utility of a high-resolution meiotic crossover map in potato

Alexandre Marand[1], Hainan Zhao[1], Courtney Leisner[2], Emily Crisovan[2], Robin Buell[2], Andy Hamernik[1], Shelley Jansky[1], Jiming Jiang[1]
[1]University Of Wisconsin-Madison, [2]Michigan State University

Genetic maps, an ordered set of markers based on relative recombination frequency, provide the necessary framework for numerous genetic analyses. Here we present an ultra dense reference sequence-based genetic map in potato, facilitated by whole genome low-coverage resequencing of 90 individuals from an F1 mapping population. A total of 770 crossovers (837 cM) were identified, with a subset of 539 crossovers mapping at high resolution (< 10kb) using 1.3 million biallelic single nucleotide polymorphisms between the homologous chromosomes of the heterozygous female parent. A traditional genetic mapping approach using the same population and raw sequencing data was implemented to validate the reference-based map and to anchor de novo assembled scaffold sequence from the male homozygous parent for future pseudomolecule construction. Additionally, our high resolution crossover map identified genomic regions displaying significant segregation distortion as a result of gametic selection, with one region in particular overlapping the self-incompatibility locus of potato. In summary, we demonstrate the versatility of reference sequenced-based genetic maps constructed by whole genome resequencing and provide the first insights into the recombination landscape of potato at a fine scale.

295

Genome-wide association study for female fertility traits in Chinese holstein population

A. Liu[1,2], Y. Wang[2], M. S. Lund[1], G. Su
[1]*Department of Molecular Biology and Genetics, Aarhus University, Tjele, Denmark*
[2]*College of Animal Science and Technology, China Agricultural University, Beijing, China*

Background: Female fertility traits are the most economically important functional traits affecting the efficiency of the dairy industry. The objective of the present study was to detect quantitative trait loci (QTL) for 12 fertility traits in Chinese Holstein population by genome-wide association study (GWAS).

Materials and methods: Data of 168 bulls and 4,555 cows from 32 herds in Beijing, China were used in this study. Traits in the analysis included age at first insemination, interval from calving to first insemination, interval from first to last insemination, conception rate for first insemination, number of inseminations per conception, non-return rates at 56 and 90 days after first insemination. The performances of heifers and cows were considered as different traits. Daughter yield deviation (DYD) derived from conventional genetic evaluation of 90,027 cows from the same population was used as pseudo phenotype. The animals were genotyped using the Illumina BovineSNP50 BeadChip. The analysis of association between SNP and pseudo phenotype was performed using a linear mixed model with a single SNP regression, applying the DMU package.

Results and conclusions: A total of 43 SNPs significant at the 5% chromosome-wise level, locating on 18 chromosomes, were associated with one or more fertility traits. Further, 13 of the 43 detected SNPs are located within the regions of known genes, and the others are less than 200 kb away from 82 known genes. QTL regions were defined so that each QTL region should contain at least 2 significant SNPs within a relatively narrow interval (less than 1.5 Mb). Consequently, 6 QTL regions locating on BTA 5, 6, 8, 14, 18, 25 were detected in this study. These regions will be helpful to identify candidate genes for female fertility traits in dairy cattle.

ICQG 5
Author Index for Posters

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

NOTES

Things to do, places to see…

Downtown Madison

Wisconsin State Capitol
http://tours.wisconsin.gov/Default.aspx
2 East Main Street
Madison, WI 53703

Overture Center for the Arts
www.overturecenter.org
201 State Street
Madison, WI 53703
608-258-4141

Madison Museum of Contemporary Art (MMoCA)
www.mmoca.org
227 State Street
Madison, WI 53703
608-257-0158

Madison Children's Museum
www.madisonchildrensmuseum.org
100 North Hamilton Street
Madison, WI 53703
608-256-6445

Wisconsin Historical Museum
www.historicalmuseum.wisconsinhistory.org
30 North Carroll Street
Madison, WI 537003
608-264-6555

Wisconsin Veterans Museum
www.wisvetmuseum.com
30 West Mifflin Street
Madison, WI 53703
608-267-1799

State Street
A one-mile pedestrian mall lined with eclectic specialty shops, restaurants and outdoor cafes. Stat Street extends from the corners of Carroll and Mifflin Streets, westward to Lake Street, adjoining the campus of UW-Madison at Library Mall.

Dane County Farmers' Market
Saturday market located on the Capitol Square – open from 6AM – 2PM
Wednesday market located on Martin Luther King Jr. Blvd. – open from 8:30AM – 2:00PM
www.dcfm.org

Concerts on the Rooftop (Monona Terrace)
June 16 – 7PM – 9PM – Doors open at 5:30PM
http://mononaterrace.com/community/all-programs/category/concerts_entertainment_events#concerts_on_the_rooftop
One John Nolen Drive
Madison, WI 53703
608-261-4000

UW-Madison Campus

Campus Walking Tours
https://info.wisc.edu/campus-tours/daily-walking-tour/
****Reservations are recommended****
Tours leave from Union South, 1308 West Dayton Street
Explore the University of Wisconsin-Madison through a guided walking tour. These 100-minute tours of campus are led by experienced tour guides and highlight campus life, academics and the history of the university. They are offered various weekdays and on the weekends. Tours leave from Union South, 1308 West Dayton Street. See website for more details and to make reservations.

Bascom Hill Historic Architecture Walking Tour
June 16 – 6PM – 7:30PM
http://www.madisonpreservation.org/#!2016-tour-info/ppop1
****Reservations are NOT necessary****
Starting Location: Library Mall entrance to the Wisconsin Historical Society headquarters on the campus end of State Street

Chazen Museum of Art
www.chazen.wisc.edu
800 University Avenue
Madison, WI 53706
608-263-2246

UW Geology Museum
www.geoscience.wisc.edu/museum_wp
1215 West Dayton Street
Madison, WI 53706
608-262-1412

Babcock Hall Dairy Store
http://babcockhalldairystore.wisc.edu/
1605 Linden Drive
Madison, WI 53706
608-262-3045

Allen Centennial Garden
http://allencentennialgarden.org/
620 Babcock Drive
Madison, WI 53706

Memorial Union
https://union.wisc.edu/
800 Langdon Street
Madison, WI 53715
608-890-3000

Third Thursdays at the Pyle Center
June 16 – 4PM – 7PM
http://conferencing.uwex.edu/thirdthursdays.cfm
702 Langdon Street
Madison, WI 53715
608-890-0392

Lakeshore Path and Picnic Point
http://lakeshorepreserve.wisc.edu/

Places Outside of Downtown and Campus

UW-Madison Arboretum
https://arboretum.wisc.edu/
1207 Seminole Highway
Madison, WI 53711
608-263-7888

Wingra Boats
www.wingraboats.com
824 Knickerbocker Street
Madison, WI 53717
608-233-5332

Rutabaga Rentals and Paddlesports
http://www.rutabaga.com/rutabaga-rentals
220 West Broadway
Madison, WI 53716
608-223-9300

Madison BCycle (Experience Madison by bike)
https://madison.bcycle.com/
40 stations. 350 bikes

Henry Vilas Zoo
www.vilaszoo.org
702 South Randall Avenue
Madison, WI 53715
608-266-4732

Capital Brewery
www.capitalbrewery.com
7734 Terrace Avenue
Middleton, WI 53562
608-836-7100

Olbrich Botanical Gardens
www.olbrich.org
3330 Atwood Avenue
Madison, WI 53706
608-246-4550

Madison Restaurant Information

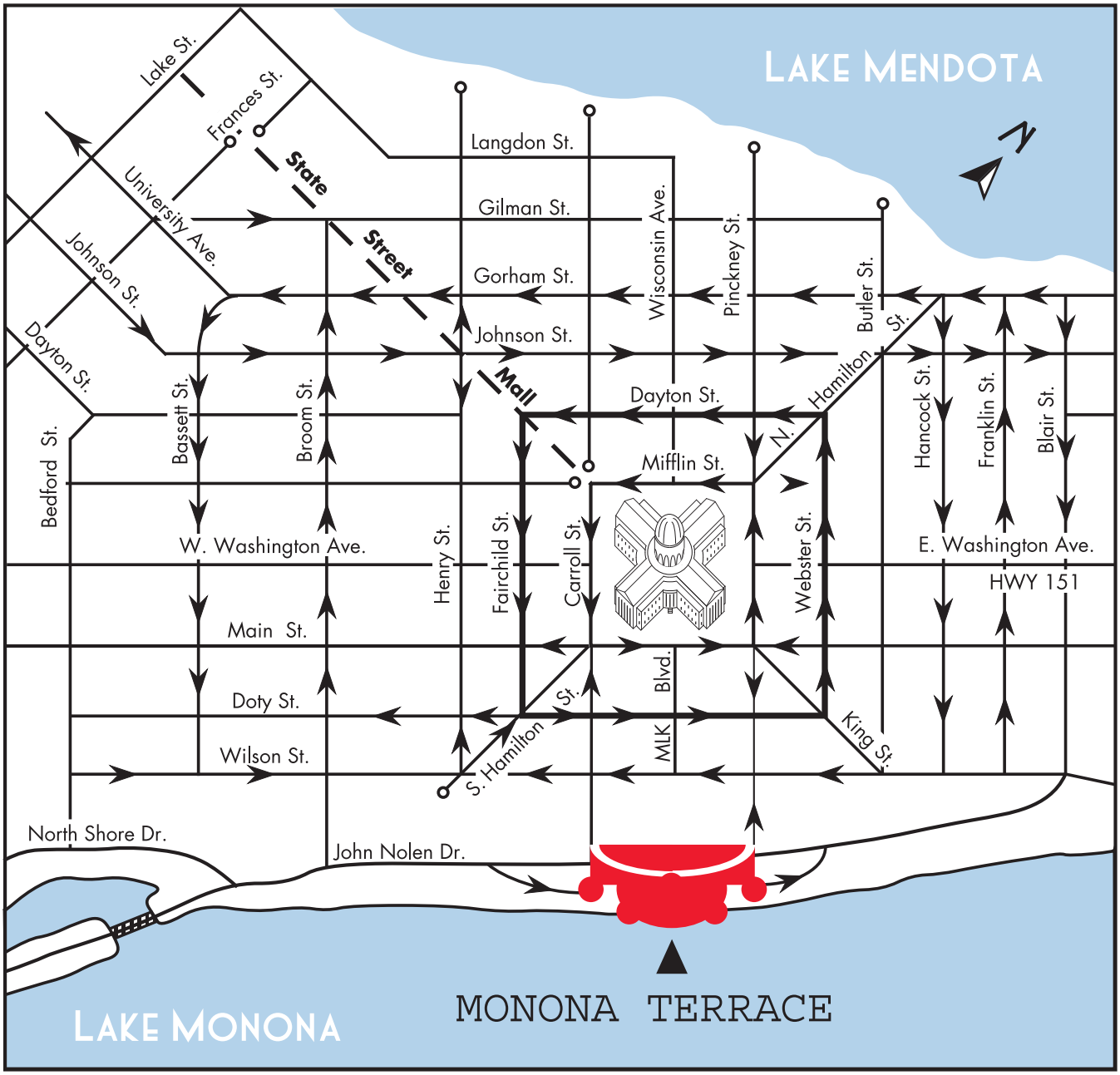
http://www.visitmadison.com/restaurants/

http://www.visitmadison.com/restaurants/wisconsin-culinary-delights/

http://www.foxnews.com/leisure/2013/09/13/top-10-reasons-why-madison-is-foodie-paradise/

Map of the Madison Isthmus area.

Map of downtown/capitol area.



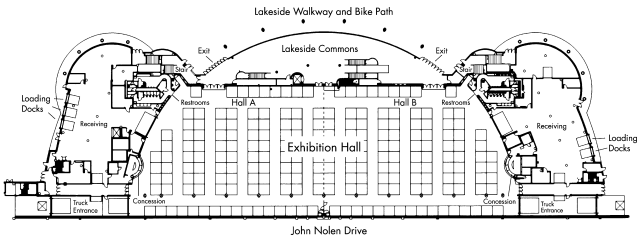
WHERE ARE WE?

All events will be held at the Monona Terrace Community and Convention Center.

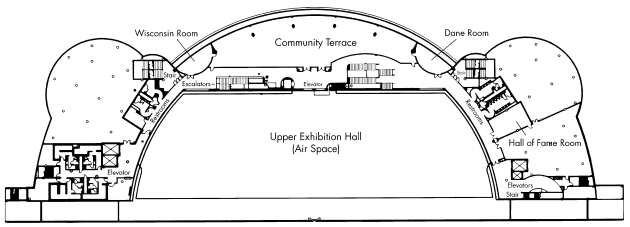
The keynote presentation will be held in Ballroom AB with the welcome reception following in the Grand Terrace. All other general sessions will be held in Exhibit Hall B. The exhibitors and poster sessions will be in Exhibit Hall A. The Wednesday happy hour will be held on the rooftop center circle (rain back-up is Ballroom AB and the Grand Terrace East).

MONONA TERRACE COMMUNITY AND CONVENTION CENTER

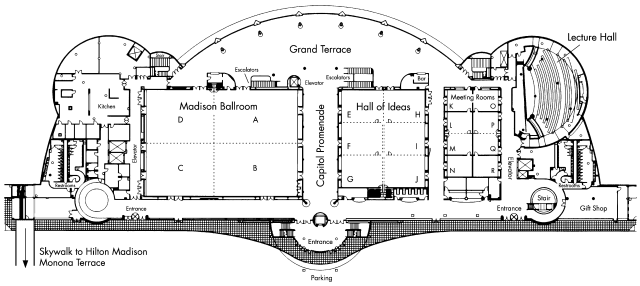
LEVEL 1 - LAKESIDE



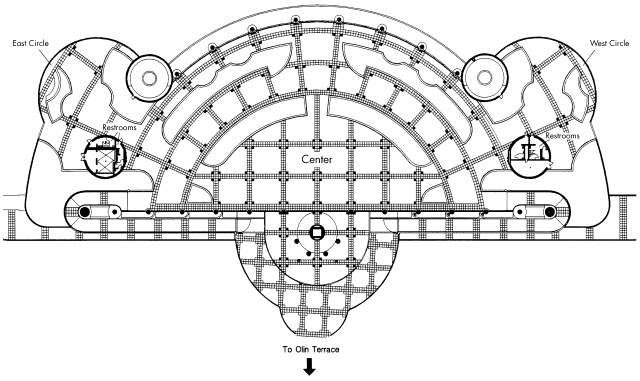
LEVEL 2 - MEZZANINE



LEVEL 4 - MEETING ROOMS / GRAND TERRACE



LEVEL 5 - ROOFTOP GARDENS



SCHEDULE AT A GLANCE

Sunday June 12

2:00 – 8:00 PM	Registration Open: Monona Terrace Community and Convention Center
6:00 – 7:00 PM	Welcome and Keynote Address – Ballroom AB
7:00 – 8:00 PM	Welcome Reception – Grand Terrace

Monday June 13

8:30AM - 11:50 PM	Session I: Quantitative Genetics in the Genomics Era – Exhibit Hall B
10:00 – 10:30 AM	Coffee break – Exhibit Hall A
11:50 – 1:20 PM	Lunch – Exhibit Hall A
1:20 - 4:30 PM	Session II: Advances in Modeling and Computational Tools – Exhibit Hall B
3:05 – 3:35 PM	Coffee break – Exhibit Hall A
4:30 – 6:30 PM	Poster Session – Exhibit Hall A

Tuesday June 14

8:30AM - 12:05 PM	Session III: The Genetic Architecture of Quantitative Traits – Exhibit Hall B
10:00 – 10:30 AM	Coffee break – Exhibit Hall A
12:05 – 1:20 PM	Lunch – Exhibit Hall A
1:20 - 4:30 PM	Session IV: Networks and Causality in Genetics – Exhibit Hall B
3:05 – 3:35 PM	Coffee break – Exhibit Hall A
4:30 – 6:30 PM	Poster Session – Exhibit Hall A

Wednesday June 15

8:30AM - 12:05 PM	Session V: Population and Evolutionary Genetics in the Genomics Era – Exhibit Hall B
10:00 – 10:30 AM	Coffee break – Exhibit Hall A
12:05 – 4:30 PM	Afternoon social activities
5:00 - 7:00 PM	Happy Hour – Monona Terrace Rooftop Center Circle

Thursday June 16

8:30AM – 12:05 PM	Session VI: Big Data in Genomics – Exhibit Hall B
10:00 – 10:30 AM	Coffee break – Exhibit Hall A
12:05 – 1:20 PM	Lunch – Exhibit Hall A
1:20 – 4:30 PM	Session VII: Epistasis and G x E Interaction – Exhibit Hall B
3:05 – 3:35 PM	Coffee break – Exhibit Hall A
4:30 – 6:30 PM	Poster Session – Exhibit Hall A

Friday June 17

8:30AM - 12:05 PM	Session VIII: Genomic Information in Prediction – Exhibit Hall B
10:00 – 10:30 AM	Coffee break – Exhibit Hall A
12:05 – 1:20 PM	Lunch – Exhibit Hall A
1:20 – 4:00 PM	Session IX: Translational Quantitative Genomics – Exhibit Hall B
2:50 – 3:20 PM	Coffee break – Exhibit Hall A
4:00 - 4:40 PM	Closing
6:30 - 11:00 PM	Optional Conference Dinner: Overture Center for the Arts